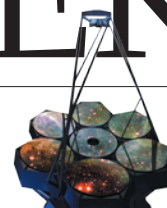


THIS WEEK

EDITORIALS

WORLD VIEW 'Why' isn't the only challenge for mental illness research **p.509**

EVOLUTION Size matters for travelling groups of sperm **p.511**



SPACE Desert telescope gets \$40-million boost from Brazil **p.512**

Safety doesn't happen by accident

To create a strong biosafety culture, information on mishaps involving deadly pathogens must be reported and shared fully and transparently.

While the US Centers for Disease Control and Prevention (CDC) was investigating an accident involving anthrax that happened at a lab on its Atlanta, Georgia, campus in June, the agency's director Thomas Frieden got a nasty surprise. Another accident, this time involving the deadly H5N1 avian influenza virus, had been discovered at a CDC laboratory six weeks previously but had not been reported at the time. Frieden was angry, and rightly so. But this was not just a one-off — biosafety experts contend that many such incidents in secure labs worldwide go unreported.

The CDC accidents raised many justified concerns, but they also led to some undue worries in the media and to political grandstanding. The risks posed by pathogens kept in high-biocontainment labs need to be kept in perspective. Many such agents are poorly transmissible, so pose mostly local threats — as well as the risk that they will be stolen and used in bioterrorism. Few are highly transmissible and able to spark epidemics of global significance.

But some pathogens do pose such risks. In July 2003, a sustained public-health effort probably stopped the SARS (severe acute respiratory syndrome) virus from causing a pandemic. But a few months later, lab accidents infected researchers in Taiwan and Singapore. And the following year, the virus was accidentally released from a lab in China and infected a researcher, then spread to her mother — who died — and a nurse. A pandemic could well have resulted.

If staff and public health are to be protected, then accidents must be reported in full, and the long-standing lack of progress here must end. As a News article on page 515 reports, many accidents are caused not by a lack of physical barriers or regulations, but by the absence of a strong biosafety culture in labs and their oversight bodies.

A key part of such a culture is timely knowledge of all accidents and their causes. That way, organizations everywhere can quickly take on board the lessons learned. The International Federation of Biosafety Associations, among others, has proposed the creation of an international system for sharing such information confidentially, but the meagre funding needed has not been forthcoming.

A confidential system would be a start, and deserves support, but it is not enough. Regulatory and oversight bodies throughout the world should require the reporting of all serious accidents and near misses in biocontainment labs, and in particular those that occur in labs with the highest biosafety levels. Timely incident reports should also be made available on public websites — as many nuclear regulators require of power plants — perhaps with an option for sharing details and more-sensitive information confidentially.

Researchers must be given incentives to report accidents. A strong biosafety culture would clearly communicate and enforce the rules of play. Negligence should be disciplined, but researchers who have accidents while acting in good faith should not be penalized unfairly. Some of the current media and political reaction to the CDC accidents and the calls for disciplinary action against the researchers involved is unhelpful and potentially unjustified. On 22 July, Michael Farrell resigned as head of the CDC's Bioterror Rapid Response and Advanced Technology Laboratory in Atlanta, and other heads may roll, too.

As one biosafety expert told *Nature*, the current criticism of the people involved means that most researchers would probably now think twice about reporting an accident. This blame game is unhelpful. What is more important, and in everyone's interests, is to prevent future accidents. And that requires full data on accidents and why they happen. ■

Fishy business

Delays in approving genetically engineered salmon may be a taste of worse to come.

The mood at AquaBounty Technologies a year ago was buoyant. Regulators had released a draft assessment of the company's genetically engineered salmon, which grow faster than normal, and found them to be environmentally benign. A few months after the assessment's comment period closed, the company began to raise more than 6,000 kilograms of salmon at its facility in Panama, in anticipation of the final approval by the US Food and Drug Administration (FDA) that would open the gates and allow the fish onto supermarket shelves.

That optimism now lies buried alongside the fish, which were culled when the approval failed to come through. The FDA says that it is still

processing the more than 35,000 public comments made in response to the draft assessment. But for AquaBounty, based in Maynard, Massachusetts, this is just the latest in a series of delays spanning nearly 20 years (see *Nature* **497**, 17–18; 2013). Many of the FDA's deliberations have taken place behind closed doors, fuelling confusion as to the cause of the setbacks, and rumours of political interference.

As the delays have dragged on, the technology used to make AquaBounty's salmon has become outdated. In the current excitement over targeted gene editing that allows researchers to modify individual genes without leaving traces of foreign DNA, AquaBounty's salmon — which contain a gene from another species — seem like a relic.

But the company's experience may hold a cautionary message. The FDA has not yet announced how it will evaluate animals engineered with gene-editing techniques. Its discussions are again occurring in private, leaving frustrated researchers to wonder whether the fruits of these technologies will meet the same fate as the beleaguered salmon. The FDA should learn from past experiences, bring these discussions before the public, and leave political considerations at the door. ■



Cause is not everything in mental illness

Welcome steps have been made in uncovering a biological basis for schizophrenia, but for many, the question of 'why' is unimportant, says **David Adam**.

The past week has been big for mental illness. As reported last week in this journal, psychiatric researchers have uncovered a spread of genetic clues to schizophrenia, potentially shedding some biochemical light on how this dreadful disease develops. At the same time, a leading US centre for research on mental-health disorders announced a record US\$650-million donation from philanthropist Ted Stanley to boost that work (see *Nature* **511**, 393; 2014).

Good news all round. And more could yet follow: genetic understanding of psychiatric disorders, together with more research on the unusual ebb and flow of circuits in the brain, promise a revolution. Researchers of brain disorders compare the current state of their science to knowledge of cancer a decade or so ago, before molecular approaches could stratify patients and select specific treatments.

The latest study on schizophrenia could be a small step forward in this march. Or it could be another false start in a field that has endured more than its fair share. Psychiatric research has yet to provide a single reliable biomarker to aid diagnosis and treatment. Self-reported symptoms and their subjective interpretations remain the basis for clinical diagnosis. Drug companies have walked away. The task of unravelling the biological pathways that drive mental illness, which are needed before drug targets can be identified, has been declared too difficult and too expensive.

Of course, some perspective is needed. Psychiatric research had a long and painful birth. Just a generation or two ago, at a time when physicists had split the atom and biologists were deciphering the structure of DNA, a common treatment for schizophrenia and other mental disorders was a metal spike hammered up through the top of the eye socket and waggled around. With such a history, a lag of a mere decade or so behind cancer research can be taken as a sign of rapid progress.

Whether or not the latest study on the genetics of schizophrenia takes that progress forward, it has already contributed to the public debate around mental illness and public understanding of the issues. It has raised and highlighted the 'C-word': cause.

I have obsessive-compulsive disorder (OCD). That used to be a secret, but in April I published a book about the condition and my experiences of it. Despite its frequent portrayal as a behavioural quirk, OCD is a vicious and debilitating mental illness, with some similarities to the experiences of schizophrenia. Simply put, people with OCD can have some of the same dark ideas, thoughts and images as someone with schizophrenia, but the person with OCD is fully aware that they generate the thoughts themselves. (The psychosis that defines schizophrenia is characterized by a lack

of such insight, and people with the condition typically attribute the intrusions to an external source.)

I now give talks about my OCD. A frequent question from the audience is one that I am still ill-prepared to answer: "What caused it?"

I don't know, and more to the point I don't care. For 20 years or so I have battled the symptoms. More recently, I sought and received treatment for those symptoms — a high daily dose of the antidepressant sertraline hydrochloride and several months' worth of weekly sessions of cognitive behavioural therapy. It seemed to work, and without anyone — psychiatrists, psychologists or me — trying to identify the cause.

Perhaps the question from others is down to simple curiosity. I tell a human story and it is natural to want to know how such stories begin. Maybe there is a degree of self-interest because people do not want to

follow the path that I did. It could be me who is unusual in not caring about a cause, but when I find out that people have cancer or heart disease or have had a stroke, the cause of their suffering is pretty far down my list of enquiries. In the past two or three years, I have met lots of other people with OCD and other mental disorders. Many of them, like me, do not know and do not seem to care about the who, the where, the why and the when of their illness. There is only how.

The other questions are not sinister. Instead, I think that they reflect an enduring mystery of mental illness. We do not know enough about the mind and the brain to build the backstory. (And as I said earlier, existing treatments do not require it.) Into this unknown creep the myths, the misunderstandings and the agendas.

In psychoanalysis, for example, as devised by Sigmund Freud, cause is everything and, sure enough, psychoanalysts usually find a subconscious cause for a mental disorder that can be conveniently addressed by — oh, psychoanalysis.

The latest schizophrenia study helps to plug that causation gap. Schizophrenia has problems in the way that it is portrayed in the wider media, but the condition does escape the worst of the trivialization that plagues other forms of mental illness such as depression and OCD. No ignorant and patronizing opinion pieces have been penned in light of these latest developments to claim (as happens with depression, for instance) that the scientists are wrong and that schizophrenia is actually all about societal context and drug-company conspiracy. It is clearly an awful illness, and it — and by extension, other mental disorders — clearly has biological roots.

To expose those roots might lead to new treatments in future, or it might not. Either way, it helps. ■

David Adam is Editorial and Column editor for *Nature* in London.
e-mail: d.adam@nature.com

I TELL A
**HUMAN
STORY**
AND IT IS NATURAL TO
WANT TO KNOW
**HOW SUCH
STORIES BEGIN.**

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/2lrx6z

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

GLACIOLOGY

Data mix-up in sea-ice record

The recent, mysterious expansion of Antarctic ice could be overestimated because of a data-analysis error, according to US scientists.

Ian Eisenman at the Scripps Institution of Oceanography in La Jolla, California, and his colleagues found the mistake when they compared two versions of satellite data on Southern Hemisphere sea ice that were calibrated differently. The incorrect calibration of one of the data sets might account for more than half of the jump in Antarctic sea-ice growth.

The finding means that either the 2007 or the 2013 report by the Intergovernmental Panel on Climate Change reflects this error, but the authors were not able to determine which one.

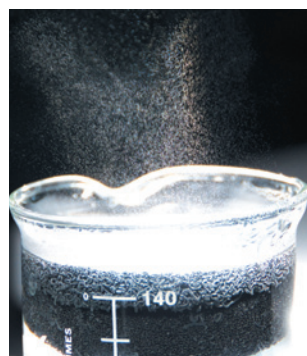
Cryosphere 8, 1289–1296 (2014)
For a longer story on this research, see go.nature.com/owasxc

MATERIALS

Sponge takes light to make steam

A sponge-like device absorbs water and solar energy to generate steam efficiently (pictured).

Gang Chen at the Massachusetts Institute of Technology in Cambridge and



his colleagues placed a layer of graphite flakes on top of a piece of carbon foam. The foam floats in water, soaking it up and wicking the liquid to the graphite, which absorbs solar radiation. Thanks to the insulating foam, heat builds up in the porous graphite layer, causing the absorbed water to evaporate.

The apparatus can trap 85% of the incoming solar energy to generate steam when sunlight is focused to ten times its normal intensity. The device could help to purify and desalinate water in remote areas, the authors say.

Nature Commun. 5, 4449 (2014)



ECOLOGY

Predictable patterns for coral-reef pest

Modelling how ocean currents spread the larvae of coral-eating starfish around Australia's Great Barrier Reef can help to identify areas that are prone to damaging epidemics of the pest.

A team led by Karlo Hock of the University of Queensland in St Lucia, Australia, used a computer model to study the distribution of larvae of the voracious crown-of-thorns starfish (*Acanthaster planci*; pictured) in the Great Barrier Reef. Areas of the reef that are

densely connected to each other through ocean currents were more likely to experience an outbreak, and to amplify it into a wider problem. The authors' model also accurately identified the specific region where epidemics most often originate.

The team suggests that careful study of reef connectivity could help to control future starfish outbreaks.

J. Appl. Ecol. <http://doi.org/tvs> (2014)

CHEMISTRY

Fighting ants make rare fluid

An ionic liquid has been observed for the first time in nature — as a mixture of venom from two rival ant species.

The tawny crazy ant *Nylanderia fulva* is displacing fire ants (*Solenopsis invicta*) in the southern United States, in part by detoxifying its enemy's venom using its own poison. Researchers led by James Davis of the University of South Alabama in Mobile show that the resulting brew

consists of ions instead of electrically neutral molecules.

The finding suggests that ionic liquids, which are commonly used in industry, also have important biological functions.

Angew. Chem. Int. Ed. <http://doi.org/f2s5zj> (2014)

GEOMORPHOLOGY

Beaches erode without storms

Sea-level rises that are unrelated to major storm events could be eroding coastlines as much as hurricanes do.

Weather and oceanographic processes that are not linked to storms cause sea levels to rise over weeks to months, but their effects have been overlooked in models of beach erosion. So Ethan Theuerkauf and his colleagues at the University of North Carolina's Institute of Marine Sciences in Morehead City studied sediment cores from six sites along Onslow Beach on the US east coast after a year of frequent sea-level changes but no major storms. They compared these cores with those obtained after a storm year and found similar levels of erosion.

The authors suggest that sea-level changes could become more frequent in this region because climate change is predicted to weaken the Gulf Stream, which can lead to these sea-level anomalies.

Geophys. Res. Lett. <http://doi.org/ttn> (2014)

IMMUNOLOGY

Inflammation on the clock

An internal clock regulates inflammation in mouse lungs.

Symptoms of some human lung diseases, including asthma, tend to vary in severity according to the time of day. Andrew Loudon and David Ray at the University of Manchester, UK, and their colleagues found that immune responses to a bacterial toxin are regulated by a circadian clock in mouse lungs. The recruitment of immune cells called neutrophils and the expression of several immune-related proteins responded rhythmically to the toxin, with neutrophil recruitment peaking at dawn.

Deleting a key 'clock gene' weakened responses to bacterial infection and reduced the effect of glucocorticoid steroids, which normally suppress inflammation. Chronic lung inflammation could be partly caused by circadian disruption, the authors say. *Nature Med.* <http://dx.doi.org/10.1038/nm.3599> (2014)

CLIMATE SCIENCE

Warming from coupled climates

Links between the climate over the North Pacific and the North Atlantic oceans could lead to abrupt climate change.

Researchers have debated whether temperature and ocean fluctuations were in sync with each other during past climate changes. Summer Praetorius and Alan Mix of Oregon State University in Corvallis studied oxygen isotopes as a proxy for ocean temperature in three sediment cores from the Gulf of Alaska covering the past 18,000 years.

By comparing the Alaska samples to cores from northern Greenland, the scientists found that climate variables such as temperature changed synchronously between about 15,500 and 11,000 years ago — shortly before the end of the last ice age.

The authors suggest that this link could have shifted heat in both oceans towards the poles at the same time, triggering abrupt climate change. They add that similar connections may be important for future warming.

Science 345, 444–448 (2014)

OPTICS

Transistor uses single photons

Two teams in Germany have built transistors that control light at the single-photon level.

Transistors that switch light instead of electrical current can enable ultra-fast computing. But making optical transistors with 'gain' — when one photon affects many others to drive further switches — has been tricky because photons do not interact with each other.

To overcome this problem, a team at the Max Planck Institute of Quantum Optics in Garching and a separate group at the University of Stuttgart passed a single photon through a cloud of ultracold rubidium atoms.

SOCIAL SELECTION

Popular articles on social media

Beef's big impact on Earth

Beef is suddenly big on social media, thanks to two recent papers investigating the global effects of livestock farming. They make the case that beef production has a bigger impact on greenhouse-gas emissions and on the use of nitrogen and water than does the production of pork and poultry, for instance. Tim Thomson, a physician and molecular biologist at the Molecular Biology Institute of Barcelona in Spain, tweeted: "Do not imitate Americans: Eat less beef and you will mitigate environmental costs of diet." But Jared Decker, a beef-cattle geneticist at the University of Missouri in Columbia, tweeted that cattle have a relatively small carbon footprint compared to other industry sectors, adding: "Wouldn't changing transportation & energy be more important?"

Clim. Change <http://doi.org/tvw> (2014); *Proc. Natl Acad. Sci. USA* <http://doi.org/tvx> (2014)



Based on data from altmetric.com. Altmetric is supported by Macmillan Science and Education, which owns Nature Publishing Group.

➔ **NATURE.COM**
For more on popular papers:
go.nature.com/ajre5g

The photon converted one atom into a type of large, excited particle called a Rydberg atom, which blocked the next photon from passing through.

In the Stuttgart team's transistor, one photon diverted another 10, whereas in the Max Planck device, a photon controlled a further 20.

Phys. Rev. Lett. 113, 053602 (2014); 113, 053601 (2014)

EVOLUTION

Sperm are speedier in groups

In the face of competition, sperm cells travel faster when

they move together in groups of an optimal size.

A team led by Heidi Fisher at Harvard University in Cambridge, Massachusetts, studied rodent sperm cells under a microscope, and used a mathematical model to analyse their swimming behaviour. They found that, in comparison to solitary sperm or those in larger groups, intermediate-sized aggregates of six or seven sperm (**pictured**) tend to migrate the fastest, by taking a more direct path.

Sperm cells from a sexually promiscuous species of deer mouse, *Peromyscus maniculatus*, were faster and more likely to form optimally sized clumps than were similarly shaped sperm from a monogamous sister species, *Peromyscus polionotus*.

The results show how sexual selection can shape the evolution of cooperation. *Proc. R. Soc. B* 281, 20140296 (2014)

➔ **NATURE.COM**
For the latest research published by Nature visit:
www.nature.com/latestresearch



SEVEN DAYS

The news in brief

RESEARCH

Malaria vaccine

London-based pharmaceutical company GlaxoSmithKline has asked the European Medicines Agency to review its malaria vaccine, which would be the world's first, under a programme designed to license medicines for use primarily outside Europe. A study published on 29 July indicates that the vaccine is 45% effective in preventing infections for 18 months after its administration in children aged 5–17 months (The RTS,S Clinical Trials Partnership *PLoS Med.* **11**, e1001685; 2014).

Agriculture body

The US Department of Agriculture announced the creation of the Foundation for Food and Agricultural Research on 23 July. Agricultural researchers have long called for such a body as a source of extra funding; it will focus on a variety of issues, including plant and animal health and nutrition. Congress has provided US\$200 million for the non-profit foundation, to be matched by external donations. The body will be guided by a board of 15 directors.

Nuclear risks

The US nuclear-power industry must be more proactive in seeking and acting on information about potential threats to nuclear plants, concludes a report released on 24 July by the US National Research Council. The report to the US Congress sought lessons from the 2011 meltdown at the Fukushima Daiichi nuclear plant in Japan. The industry, it says, should incorporate more modern risk assessments of earthquakes,

tsunamis and solar storms that could cut plants' electrical power, a key factor in the Fukushima accident.

BUSINESS

Drugs repurposed

UK scientists will gain access to experimental drugs that have been deprioritized by pharmaceutical companies, the government said on 21 July. The UK Medical Research Council (MRC) will fund projects to develop treatments from as-yet-unnamed compounds from seven companies. The programme follows a previous MRC-funded scheme to repurpose experimental drugs that AstraZeneca had suspended from development.

POLICY

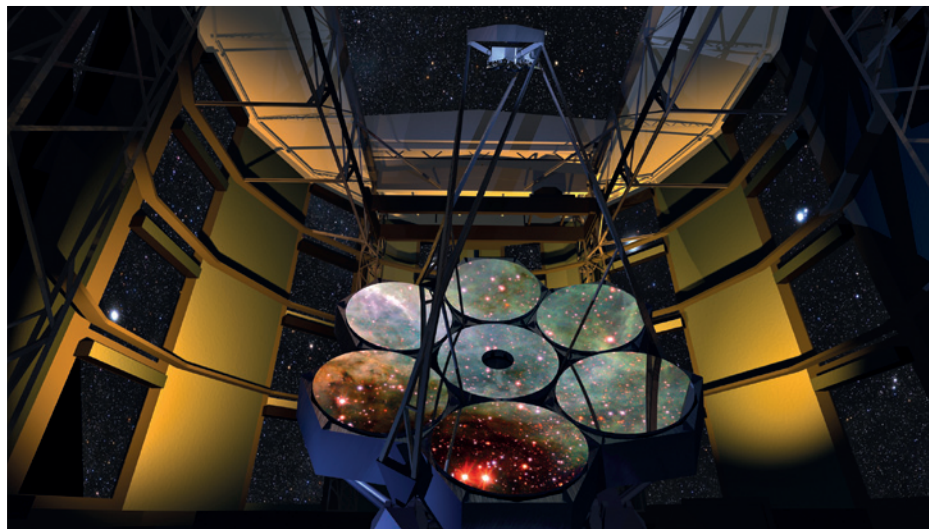
CDC safety group

In the wake of biosafety mishaps in the past two months, the US Centers for Disease Control and Prevention (CDC) has formed an external laboratory-safety working group. The group will advise the agency's head and its director of laboratory safety on corrective actions for the CDC's labs. It will also identify training and oversight needs, and review biosafety protocols. The group's first meeting will be in early August. See pages 507 and 515 for more.

Polish space agency

Poland is set to create a national space agency, even though it is already a member of the

FAPESP will contribute \$40 million to the project, opening up access to researchers from the state of São Paulo. But the foundation hopes to share the costs with the Ministry of Science and Technology of Brazil to allow astronomers from across the country to access the telescope when it begins operations in 2021. See go.nature.com/k3tsgv for more.



GIANT MAGELLAN TELESCOPE

Giant telescope gets boost from Brazil

Brazil's São Paulo Research Foundation (FAPESP) confirmed on 22 July that it will join the US\$880-million Giant Magellan Telescope (GMT), to be located at the Carnegie Institution for Science's Las Campanas Observatory in the Atacama Desert, Chile. The 25-metre instrument is one of three competing megatelescopes to be built in the next decade.

European Space Agency (ESA). On 25 July, its parliament made the decision to establish the Polish Space Agency (POLSA), which will oversee space research and — the country hopes — give Polish researchers easier access to ESA projects and make it simpler to set up space-related companies and research centres.

EVENTS

Long drive on Mars

NASA's Opportunity rover has clocked up more than 40 kilometres on Mars — breaking the record for long-distance driving on an extraterrestrial world. On 28 July, the space agency announced that Opportunity had surpassed the 1973 record

held by the Soviet Union's Lunokhod 2 Moon rover. Experts had been unsure exactly how far the Soviet rover had travelled (see *Nature* **498**, 284–285; 2013), but new calculations involving satellite imagery of its tracks show that it went about 39 kilometres. Opportunity landed on Mars in January 2004 and has been driving ever since.

Ebola outbreak

The largest recorded outbreak of Ebola virus has spread to Nigeria, which has reported its first case of the disease: an aeroplane passenger died in Lagos on 25 July after travelling from Liberia. The World Health Organization says that 672 people have so far died in the outbreak, which is concentrated in the West African countries of Liberia, Sierra Leone and Guinea. See page 520 for more.

PEOPLE

Scripps resignation

Michael Marletta (pictured) is stepping down as president of the Scripps Research Institute in La Jolla, California, according to a 21 July statement from the institute. Marletta's plan for a US\$600-million merger between Scripps and the University of Southern California in Los Angeles drew faculty ire and a vote of no confidence this month. Scripps



faces a \$21-million budget deficit. See go.nature.com/cvozm for more.

Lab chief resigns

The head of a laboratory in which employees may have been accidentally exposed to live anthrax has resigned. Michael Farrell stepped down as director of the Bioterror Rapid Response and Advanced Technology lab in Atlanta, Georgia, on 22 July. The lab, part of the US Centers for Disease Control and Prevention, had failed to inactivate anthrax spores properly before sending them to labs with lower biocontainment levels. Farrell had already been reassigned from his post following the incident, which happened in June.

FACILITIES

Rocket shortfall

NASA is short of money to carry out the first flight test of its heavy-lift rocket planned for December 2017, according

to a report released on 23 July by the US Government Accountability Office. The space agency expects to spend nearly US\$12 billion developing the vehicles for the first launch; the report says it would need some \$400 million more. NASA engineers are gearing up to test parts of the rocket engines — modified from the space-shuttle programme — at a centre in Mississippi.

Telescope reprieve

On 21 July NASA announced a two-year reprieve for the Spitzer Space Telescope, which had been scheduled for shutdown following a review of the agency's astrophysics mission priorities in May. Spitzer observes planets, stars and galaxies at infrared wavelengths. It launched in 2003 and, after running out of coolant in 2009, began observing at slightly higher and less-optimal temperatures. Because of its Earth-trailing orbit, it can observe objects that other telescopes cannot, such as small asteroids following the planet.

Hawaii green light

Hawaii's planned Thirty Meter Telescope has cleared its last major legal hurdle, and construction can now begin on Mauna Kea. The state's Board of Land and Natural Resources gave approval for the telescope

COMING UP

31 JULY

The New York Academy of Sciences holds a symposium in New York city to honour the work of Marshall Nirenberg, who was awarded a Nobel prize in 1968 for his part in deciphering the genetic code and protein synthesis. One session will see experts discuss legal, ethical and social issues related to applications of the genetic code.

go.nature.com/wq1ong

observatory to sublease the site from the University of Hawaii on 25 July. The Office of Hawaiian Affairs, which advocates for the interests of native Hawaiians, who consider the mountain holy, had contested the arrangement in a petition filed on 7 July. It later withdrew the petition.

AWARDS

Blavatnik awards

The three winners of this year's US Blavatnik National Awards were announced on 28 July. Neurobiologist Rachel Wilson from Harvard University in Cambridge, Massachusetts, was recognized for mapping the circuitry of fruit-fly brains. Adam Cohen, also from Harvard, won for his advances in imaging neural activity in real time. Marin Soljačić of the Massachusetts Institute of Technology in Cambridge was rewarded for his work on electromagnetic phenomena, including wireless battery charging. Each person receives US\$250,000 — the largest unrestricted cash prize for early-career scientists. The prizes are awarded annually by the Blavatnik Family Foundation and the New York Academy of Sciences.

► NATURE.COM

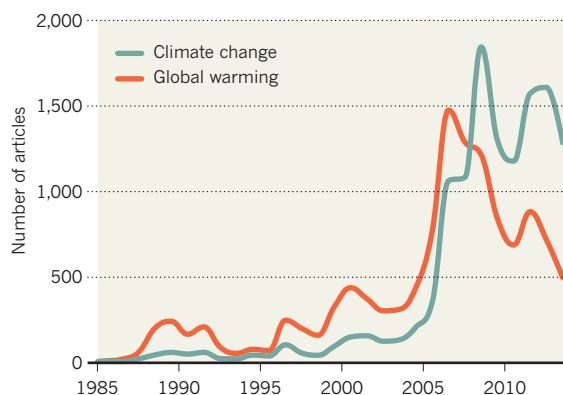
For daily news updates see:
www.nature.com/news

TREND WATCH

The New York Times now writes about 'climate change' much more than it does about 'global warming' — a move that may reflect a shift in reporting emphasis away from temperature and towards effects such as rising sea levels and ocean acidification. Roger Pielke Jr, a science-policy expert at the University of Colorado Boulder, spotted the trend using the newspaper's Chronicle tool, released for public use on 23 July. The 2009 peak for 'climate change' coincides with the United Nations climate summit in Copenhagen.

CLIMATE CHANGE IN THE MEDIA

A publicly accessible online tool shows how use of the terms 'climate change' and 'global warming' has varied in *The New York Times*.



NEWS IN FOCUS

GEOPHYSICS Drilling project prepares to drop sensors into heart of earthquake **p.516**

CELL CULTURE *Mycoplasma* bacteria contaminate 10% of cell experiments **p.518**

INFECTIOUS DISEASES Funding, logistics and inertia hold up Ebola treatments **p.520**



LIQUID BIOPSIES Blood samples hold promise for detecting cancer **p.524**

TODD PARKER/CDC



An 'inactivated' sample of the anthrax pathogen *Bacillus anthracis* began growing at a US laboratory.

INFECTIOUS DISEASES

Biosafety controls come under fire

Experts call for a stronger safety culture at secure sites after incidents involving anthrax and flu in a US laboratory.

BY DECLAN BUTLER

Recent accidents involving deadly pathogens at a leading laboratory in the United States highlight the need for a major global rethink of biosafety controls, experts say.

The Centers for Disease Control and Prevention (CDC) in Atlanta, Georgia, reported two accidents involving anthrax and the deadly H5N1 influenza virus. Biosafety professionals argue that such incidents show that without a strong culture of biosafety, even highly secure facilities are susceptible to errors that could place workers and the public at risk.

Until now, biosafety has mostly been about physical biocontainment, meeting safety regulations and following recognized standard operating procedures, says Tim Trevan, who is executive director of the International Council for the Life Sciences, a non-profit body in McLean, Virginia, that advises on biosafety policies. But organizations also need to focus on developing a stronger safety ethos, he says. "I hope that the accidents will trigger profound cultural change, not just at the CDC but at high-containment labs everywhere."

The incidents at the CDC occurred in March and June. In the first, a sample of a

low-virulence flu virus that was transferred to another laboratory had been accidentally contaminated with the lethal H5N1 avian flu strain. The second incident involved the transfer of potentially inadequately inactivated anthrax bacteria from a biosafety-level-3 laboratory to a lab with a lower safety level that was not equipped to handle such a dangerous pathogen.

The events have triggered a media and political storm, leading to considerable pressure on the CDC and other US labs to improve their practices. On 16 July, Thomas Frieden, the CDC director, was called to testify on the anthrax incident before a House of Representatives committee. "The fact that something like this could happen in such a superb laboratory is unsettling because it tells me that we need to look at our culture of safety throughout all of our laboratories," he said ahead of the hearing. "We are definitely looking at the implications for laboratories around the country and around the world."

Last week, the CDC announced the creation of an independent committee to review the agency's safeguards. Safety culture is among the topics the committee will discuss when it meets for the first time next month.

The term 'culture of safety' is more than just jargon — management frameworks to aid organizational safety are well established in, for example, the airline and nuclear-power industries, says Trevan. Creating such a culture requires practices and training that are targeted at addressing risks in a structured manner, and constantly monitoring and improving performance. Yet researchers and oversight bodies all too often have a "checkbox culture", he adds.

This can result in a management mentality of "we don't care if the plan works, as long as you have a plan", says Sean Kaufman, president of Behavioral-Based Improvement Solutions — a company in Woodstock, Georgia, that trains staff who work in biocontainment laboratories. He says that institutions are often reluctant to spend resources on improving practices: "Typically, leadership will only invest so much in biosafety; the bare minimum required to keep them out of trouble and in compliance."

Over the past decade, more attention has been paid to biosafety culture as the field has become increasingly professional. In 2008, the European Committee for Standardization (CEN) in Brussels adopted the first internationally recognized management framework for ►

► organizational safety in facilities handling dangerous pathogens: CEN Workshop Agreement (CWA) 15793. This voluntary framework is currently being adapted to become an International Organization for Standardization (ISO) standard, which would give it worldwide recognition.

The World Health Organization (WHO) recommends that organizations adopt CWA 15793, says Nicoletta Previsani, who was a former head of biosafety and laboratory biosecurity at the WHO and is now responsible for containment at its polio-eradication programme. “CWA 15793 really is a major shift in thinking,” she says, adding that its implementation nonetheless requires considerable investment.

The WHO has adopted the standard for oversight of the two laboratories holding the last stocks of the smallpox virus — one at the CDC in Atlanta, and the other near Novosibirsk in Russia. It has also specifically recommended that facilities carrying out risky gain-of-function flu research, which increases the transmissibility, virulence or host range of viruses, be CWA compliant or equivalent.

But wider uptake has so far been limited. The CDC, for example, has not fully implemented the standard, and in a survey last year of 118 members of the European Biosafety Association, three-quarters of whom were biosafety professionals, just 33% reported that they were using CWA 15793 in their institutions, and 15% had never even heard of it. Reasons given for not implementing the standard included a lack of resources, its “excessive” nature and the availability of similar national standards.

However, many organizations are using the CWA standard to improve biosafety management without going to the time and expense of seeking formal certification, says Gary Burns, a UK biosafety consultant who was vice-chair of the group that developed CWA 15793. He hopes that if it is adopted as an ISO standard, this will lead to greater formal and informal use.

But such safety-management standards are not a “magic bullet”, cautions Maureen Ellis, executive director of the International Federation of Biosafety Associations (IFBA) in Ottawa, Canada, because to be effective, all staff must buy into them. Researchers too often consider biosafety as an extra burden, “something they have to do because the rules say so”, she explains.

The IFBA has sought funding to advocate for improved biosafety cultures in laboratories, but funders are not interested, Ellis adds, partly because tangible outcomes are difficult to measure. “There is money for diagnostics and research, but ask for money for biosafety and it’s just not there, as it is lower priority,” she says. ■



Researchers are set to drill a 1.3-kilometre borehole in a seismic fault near Whataroa, New Zealand.

GEOPHYSICS

Project drills deep into coming quake

Sensors in borehole at New Zealand seismic fault will peek under the surface of impending rupture.

BY KATIA MOSKVITCH

For the first time, researchers are preparing to drop a battery of sensors deep into a seismic fault to record the build-up and occurrence of a massive earthquake.

An international team will drill a 1.3-kilometre hole in the Alpine Fault in New Zealand, through which they will gather crucial

data that could help to predict future quakes. The fault ruptures roughly every 330 years, triggering a quake of up to magnitude 8 (K. R. Berryman *et al. Science* **336**, 1690–1693; 2012). The most recent earthquake was in 1717, so the next one is expected any time now.

“If we go on to record the next earthquake, then our experiment will be very, very special,” says Rupert Sutherland, a tectonic geologist at

SOURCE: GNS SCIENCE
GNS Science, a government-run Earth-science organization in Lower Hutt, New Zealand, and one of the project's leaders. "A complete record of events leading up to and during a large earthquake could provide a basis for earthquake forecasting in other geological faults."

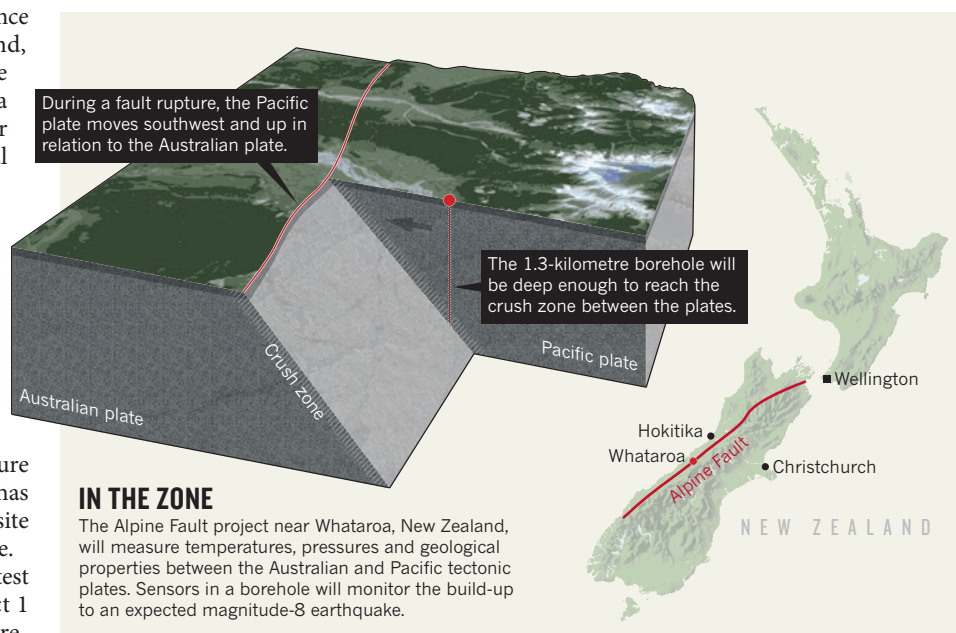
The Alpine Fault, which runs for about 600 kilometres along the west coast of South Island, marks the boundary between the Pacific and Australian plates (see 'In the zone'). Every year, these plates slide past each other by about 2.5 centimetres, building up pressure. Geologists are confident that the fault is "ready to break in its next earthquake", says Sutherland — with a 28% chance of a rupture in the coming 50 years. The Alpine Fault has been specifically selected for the drilling site because it is so late in this earthquake cycle.

In 2011, Sutherland's team completed a test phase — the Deep Fault Drilling Project 1 (DFDP-1) — in which they sank two boreholes, the biggest reaching 151 metres into the fault. In the next two weeks, work is beginning on the DFDP-2, which will drill a hole 10 centimetres across and 1.3 kilometres deep at the same site near the village of Whataroa. At that depth, the team will reach the 'crush zone' where the two plates meet, and will be able to take measurements representative of conditions deep in the crust, where earthquakes originate.

The DFDP-2 project will cost about US\$2 million, and is being backed by the International Continental Scientific Drilling Program in Potsdam, Germany, and the Marsden Fund of the Royal Society of New Zealand in Wellington.

The first part of the experiment will involve collecting geological samples and inserting sensors into a shallow borehole to measure the temperature and pressure inside the fault. The hole will then be reinforced and deepened before instruments that are able to record key indicators of seismic activity — including images, sound, temperature and pressure — are lowered into the fault. The team hopes to complete all drilling and sensor-laying work at the borehole by early December.

Data collected by the sensors will be fed into computer simulations to test theories of how faults rupture, and will help the team to develop detailed models of how the fault



IN THE ZONE

The Alpine Fault project near Whataroa, New Zealand, will measure temperatures, pressures and geological properties between the Australian and Pacific tectonic plates. Sensors in a borehole will monitor the build-up to an expected magnitude-8 earthquake.

behaves at different points in the earthquake cycle. One idea that will be tested is that large differences in groundwater pressures on either side of the fault zone could indicate whether a quake is imminent.

"The fault appears to currently form an impermeable barrier, and it's likely that time-dependent differences in groundwater pressure on either side of the fault play a role in governing earthquake nucleation processes and the radiation of seismic waves," says John

Townend, a seismologist at Victoria University of Wellington, who is part of the project.

"If we go on to record the next earthquake, then our experiment will be very, very special."

The work will also help to improve understanding of plate-boundary mechanics and seismic hazards, says David Boon, a geologist at the British Geological Survey in Cardiff, who is not involved in the research. "This drilling will underpin the science of modelling stress build-up in the crust and, importantly, stress release — which can cause large destructive earthquakes, and secondary hazards such as tsunamis, landslides and liquefaction", in which soil behaves like liquid.

Researchers have used deep-drilling

data for models before — but only in the aftermath of a quake. "To develop computer simulations for how earthquakes happen, information about initial conditions within the geological fault that is being ruptured is essential," says Sutherland. "After our experiment, realistic data — based on observation — can be used to construct models, so they will have much more value."

The only previous major attempt to peek inside an active fault is the San Andreas Fault Observatory at Depth (SAFOD), a 3.2-kilometre borehole near Parkfield, California. But this was done at a 'creeping' section of the fault, which experiences regular but small quakes rather than infrequent large ruptures.

Although not as deep as SAFOD, the DFDP-2 is calling on the US project's experience, says Cliff Thurber, a seismologist at the University of Wisconsin–Madison and a participant in the DFDP-2. In particular, the team hopes to learn from SAFOD's technical setbacks, which included instruments breaking because of the enormous heat and pressures deep underground. Thurber is also already thinking about a next borehole project, which he would like to see get even closer to an earthquake epicentre. "DFDP-2 is a great project, but my hope would be that there will be a DFDP-3 to reach deeper sometime soon," he says. ■



TOP STORY



Flaws emerge in tree of life built from RNA
go.nature.com/aaak8a

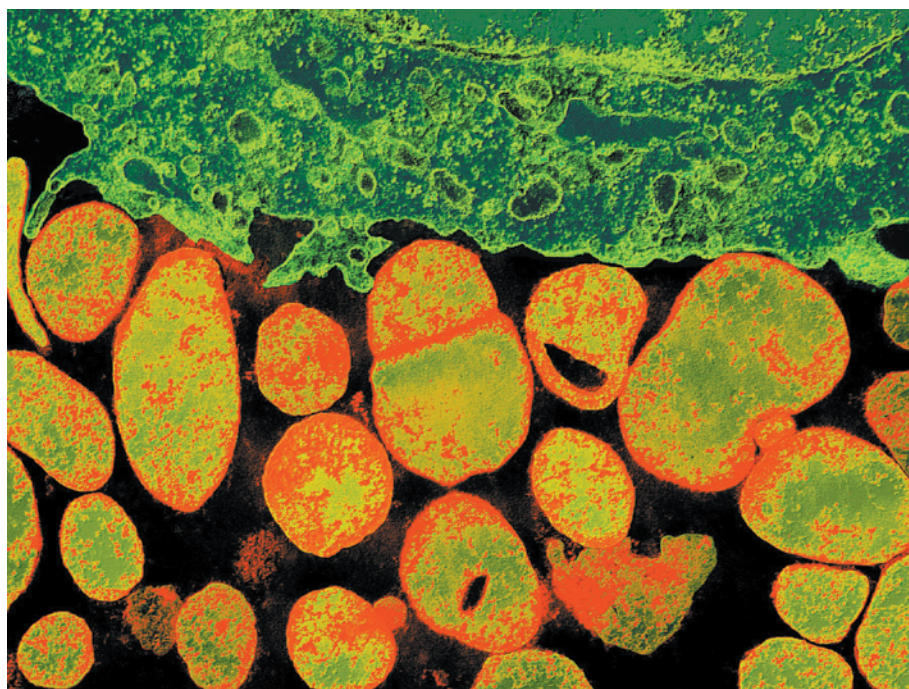
MORE NEWS

- Asteroid almost didn't wipe out the dinosaurs go.nature.com/pndxyb
- Southwestern US depletes its groundwater reservoirs go.nature.com/cowany
- Researchers call for easy access to contraception go.nature.com/xpgtfc

NATURE PODCAST



Replaying evolution in the lab, dam-busters in the United States and Earth's violent early years nature.com/nature/podcast



Mycoplasma (orange) can surround and infect cells and invalidate results of gene-expression studies.

CELL CULTURE

Contamination hits cell work

Mycoplasma infestations are widespread and costing laboratories millions of dollars in lost research.

BY EWEN CALLAWAY

John Hogenesch saw the anguished look on his technician's face and knew instantly that her experiments had gone haywire; he also had a pretty good idea of why. "Check your culture for *Mycoplasma* contamination," he advised. The bacterium is notorious for infecting cell cultures, and had indeed compromised her experiments.

In fact, the problem is widespread. Hogenesch, a genome biologist at the University of Pennsylvania in Philadelphia and his colleague Anthony Olarerin-George have found that more than one-tenth of gene-expression studies, many published in leading journals, show evidence of *Mycoplasma* contamination¹. The infestations are undermining research findings and wasting huge amounts of money, Hogenesch says.

He should know. His lab quickly overcame an infestation last year, but a previous plague cost it some US\$100,000 and a year of research. *Mycoplasma* takes hold quickly, he

says. "All it takes is one person not to check, and — bam — you have it." The bacterium often comes from lab workers, and is not killed by the antibiotics typically used to rid cell cultures of contaminants. And unlike many other microorganisms, which turn the growth medium turbid, *Mycoplasma* leaves no visible signs of its presence.

Mycoplasma is a long-standing problem. The bacterium plagued early cultures of HeLa cells, a widely used human cell line established in the 1950s. Surveys of individual collections have long served as a warning: a 1993 study², for instance, found *Mycoplasma* in 15% of 20,000 cultures from US Food and Drug Administration labs. Companies and centres that distribute cells and reagents are now fastidious in screening for it.

To get a global perspective on the problem, Hogenesch and Olarerin-George looked for stretches of *Mycoplasma* DNA in RNA-sequence data from more than 9,000 samples — collected during experiments done between 2012 and 2013 to measure gene expression in

cultured mammalian cells. A total of 11% of the samples were found to contain *Mycoplasma* DNA at levels indicative of contamination.

Some of the studies with the highest levels were published in leading journals such as *Cell*, *Nature* and *Proceedings of the National Academy of Sciences*, says Hogenesch. Contamination does not necessarily invalidate those findings, but the bacterium can influence the expression of hundreds of genes and hinders cell growth by competing for nutrients. In one particularly contaminated data set, of individual lymphoma cells, Hogenesch and Olarerin-George identified 61 genes whose levels were altered by *Mycoplasma*. Their work is available on the bioRxiv preprint server¹, and they plan to submit their findings to a peer-reviewed journal soon.

If more than one-tenth of cell cultures are contaminated, the costs in wasted time and resources, such as repeating experiments and replacing cells, could run into hundreds of millions of dollars, says Hogenesch. In 2013, for example, the US National Institutes of Health spent about \$3 billion on research that uses cell lines, and Hogenesch estimates that about one-third of his lab costs go on tissue culture.

CONTAMINATION CRACKDOWN

Contamination is a result of "sloppy cell-culture work", says Hans Drexler, a physician-scientist at the German Collection of Microorganisms and Cell Cultures in Braunschweig. "*Mycoplasma* don't fall from the sky," he says. "They are introduced into the cell culture by people." He also is not surprised that so many cell cultures are contaminated. A similar percentage of the human leukaemia and lymphoma cell lines his lab received from other researchers between 2010 and 2013 tested positive for *Mycoplasma*, he says. But this is an improvement: one-quarter of such cultures were tainted in the early 1990s, he found³.

Drexler believes that *Mycoplasma* contamination persists because of a "black market" in cell lines — researchers often share cultures in violation of materials-transfer agreements. He estimates that 10% of the cell lines he receives from these sources are contaminated, compared with none from official suppliers. To solve the problem, he urges labs to spend the extra money on cell lines from reputable sources, and test those they have for contamination.

"There's no magic twenty-first-century bullet that's going to kill these things," Hogenesch says. "We have to be continuously vigilant, clean up the cultures that have them, and destroy the bacteria altogether." ■

1. Olarerin-George, A. O. & Hogenesch, J. B. Preprint at bioRxiv <http://dx.doi.org/10.1101/007054> (2014).
2. Rottem, S. & Barile, M. F. *Trends Biotechnol.* **11**, 143–151 (1993).
3. Drexler, H. G., Uphoff, C. C., Dirks, W. G. & MacLeod, R. A. F. *Leuk. Res.* **26**, 329–333 (2002).

MEDICAL RESEARCH

FDA debates trial-data secrecy

US drug regulator weighs up merits of disclosing preliminary results.

BY HEIDI LEDFORD

Despite a trend towards increased transparency in clinical-trial data, the US Food and Drug Administration (FDA) is asking whether there are times when participants and researchers should be kept in the dark. As pharmaceutical companies push for studies that first justify a drug's approval, then monitor safety once it reaches the market, the agency fears that publicizing the early data could bias the final results.

In raising the matter, the FDA could energize the debate about a long-standing clinical conundrum, says Iain Chalmers, coordinator of the James Lind Initiative, a group based in Oxford, UK, that aims to improve clinical trials. "There hasn't been much discussion about this," he says. "There needs to be much more."

On 11 August, the FDA will hold a public hearing in Silver Spring, Maryland, to discuss situations in which preliminary results from clinical trials should be kept confidential. The FDA is obliged to release a summary of the data that it uses to approve a drug. But the public rarely sees the data given to safety committees to decide whether a trial should continue. Even if those data are not definitive but lean one way or another, making them public may spook study participants or bias investigators towards a particular outcome, the agency fears.

SCIENCE OVER SUBJECTS

The practice of confidentiality has been debated by researchers and ethicists for some time, and FDA memos arguing in favour of withholding some interim data have not convinced everyone. Although the memos have pointed out many possible negative consequences for a trial if such data are divulged, they have ignored the ethical ramifications of keeping information from participants, says Michael Carome, director of the health-research group at Public Citizen, a non-profit consumer advocacy group in Washington DC. "The agency wants to get an answer to scientific questions," he says. "The question is: in order to get their wish, are they perhaps putting human subjects at risk?"

Changes to the drug-approval process have made that question trickier than ever. In the past, sponsors rarely submitted interim data to the FDA. But in 2007, researchers found signs that a popular diabetes drug may have been increasing deaths due to cardiovascular events such as heart attacks or strokes. As a result, the agency began to demand large, prolonged safety trials of some diabetes drugs (see

MUM'S THE WORD

When approving the diabetes medication alogliptin, the US Food and Drug Administration deviated from its usual practice of publicly releasing the supporting data.



4 JANUARY 2008 The Japanese firm Takeda Pharmaceuticals applies to the FDA for approval of its diabetes drug alogliptin.

17 DECEMBER 2008 The FDA issues detailed standards for cardiovascular-safety tests of new diabetes drugs.

28 AUGUST 2009 Researchers launch clinical trial to test cardiovascular safety of alogliptin.

25 JULY 2011 The FDA receives interim data from safety trial.

25 JANUARY 2013 The FDA approves alogliptin but does not release interim data.

3 OCTOBER 2013 Final results of safety trial are published and show no evidence of cardiovascular risk.

11 AUGUST 2014 The FDA is holding a public meeting to discuss disclosure of interim clinical-trial data.

'Mum's the word'). Pharmaceutical companies are typically allowed to market those drugs once they have showed that it raises the risk of cardiovascular events by no more than 80% relative to the control group; then they are asked to do a post-approval study demonstrating that the drug boosts that risk by no more than 30%.

Increasingly, companies are petitioning to combine the two studies into one large trial, and use interim data to clear the first hurdle.

One such case came to the FDA in 2011, when it evaluated alogliptin, a diabetes drug made by Takeda Pharmaceuticals in Osaka, Japan. Interim analyses of the drug's heart risks were conducted once 81 cardiovascular events had been seen in study participants, says William White, a specialist in preventive cardiology at the University of Connecticut School of Medicine in Farmington, who led the study. Those data showed that the drug did not greatly affect the rate of cardiovascular events, White says, but could have been misinterpreted to suggest that the drug actually lowered the risk.

EARLY RELEASE

Rather than releasing those data when it approved alogliptin in January 2013, the FDA simply announced that the findings showed that the trial was safe enough to proceed. In a March 2013 memo, Mary Parks, head of endocrinology products at the FDA's Center for Drug Evaluation and Research in Silver Spring, argued that the secrecy was necessary so that long-term safety data could be obtained in a timely fashion.

White, who says that even he did not see the data until the study was finished, says that secrecy was key to successful completion of the trial because investigators might have refused to put patients on placebo had they seen the interim data.

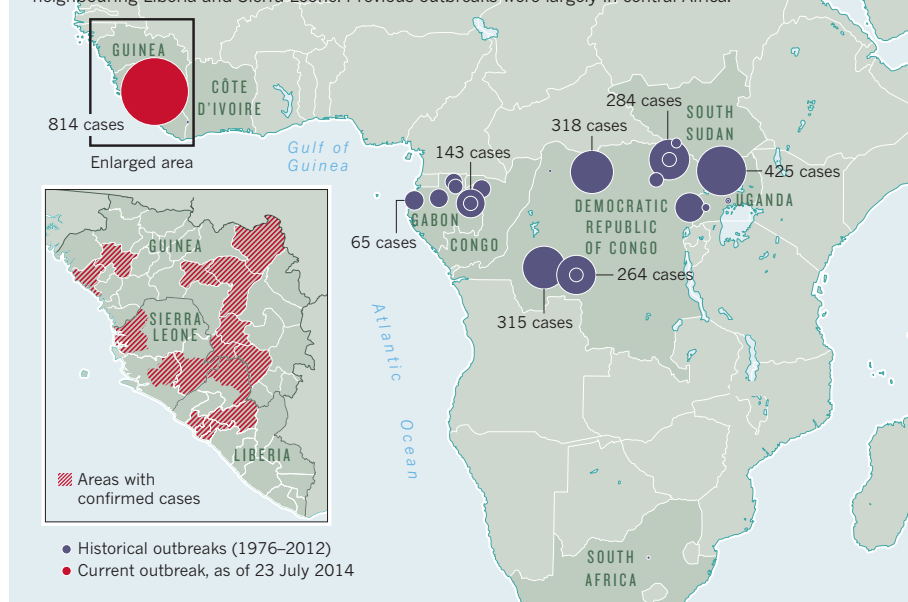
The final results, published in October, confirmed that the drug had no significant effect on cardiovascular risk (W. White *et al.* *N. Engl. J. Med.* **369**, 1327–1335; 2013).

The release of interim results might also prompt patients to abandon a trial, but that should be their choice, says Richard Lilford, chair of public health at the University of Warwick, UK. He argues that trial designers too often default to secrecy, and risk sacrificing their obligation to participants in the process. Instead, he advocates that the data be shared from the start. "Among the trial fraternity, this idea is terribly unpopular," he says. "They think that clinical trials must run until they've got a clear answer."

Paul Armstrong, a cardiologist at the University of Alberta in Canada, has served on more than 30 safety boards and says that it is standard to keep interim data confidential. But sometimes, he says, boards do decide that the benefits of revealing the data outweigh the risks. "We always ask ourselves, 'could we go and get consent for the next patient and feel confident they were adequately informed about participating in the trial?'. That is the bottom line." ■

MAJOR OUTBREAKS

With more than 800 confirmed cases so far, the current Ebola virus outbreak is the largest in recorded history. After the first cases were reported in Guinea in March, the virus spread to neighbouring Liberia and Sierra Leone. Previous outbreaks were largely in central Africa.



easy to do and not clear who you'd test it on."

The VSV vaccine seems to be a promising option because it can be used either preventively or just after a person is infected. In 2009, it was used on a German lab technician who had accidentally pricked herself with a needle carrying Ebola virus. Although it is unclear whether she was ever infected, the technician survived and suffered no ill effects from the vaccine. "Everybody in my lab would volunteer to take the vaccine," says Thomas Geisbert, a microbiologist at University of Texas Medical Branch in Galveston who is also working on the medicine.

The NIAID Vaccine Research Center in Bethesda, Maryland, has developed a vaccine that is carried by a chimpanzee adenovirus, similar to the virus that causes the common cold. The institute hopes to begin testing in healthy people as early as September. Barney Graham, deputy director of the research centre, says that the institute is talking with the Food and Drug Administration (FDA) to speed up the approval process, a position that is strengthened by the outbreak in West Africa.

Biotechnology companies are also developing treatments at a pace that could be accelerated. Mapp Biopharmaceutical in San Diego, California, is testing combinations of monoclonal antibodies that target the virus, and also hopes to begin human trials soon. And with US\$140 million from the US Department of Defense, Tekmira in Burnaby, Canada, is testing a treatment called TKM-Ebola, which uses small RNA molecules to bind the virus and target it for destruction. The company began testing the vaccine in humans in January, but on 3 July, the FDA put the study on hold until the company could provide more data on how the treatment works. Tekmira says that it is confident it will be able to restart the trial soon.

The timing of the outbreak is "unfortunate", says Armand Sprecher, a public-health specialist at Médecins Sans Frontières (also known as Doctors Without Borders) in Brussels. "If this had happened a year or two from now maybe we'd be in a better position."

A treatment could be approved by the FDA on a 'compassionate use' basis, but that process would have to mesh with a host country's rules. "A country has to request these things; it's not something we can force on them," says Gene Olinger, a virologist at the contract research organization MRIGlobal in Frederick, Maryland. "We have to follow their internal policies for drug development and for testing." ■

CORRECTION

In the News Feature 'Hello, Governor' (*Nature* **511**, 402–404; 2014), the number of co-authors for the consensus statement was given as 13 instead of 14. And the title of the report omitted the word 'scientific'. Finally, it was Governor Brown, not Elizabeth Hadly, who delivered the report to political leaders.

INFECTIOUS DISEASES

Ebola treatments caught in limbo

Logistics and lack of funds keep experimental drugs and vaccines from being used in Africa outbreak.

BY SARA REARDON

Medical relief workers fighting a burgeoning Ebola outbreak in West Africa have not been welcomed with open arms. Death was all that the hazmat-suited visitors seemed to bring. Most patients who entered the makeshift hospitals died, their families forbidden to handle their bodies. Rumours flew that these newcomers were harvesting organs and conducting fatal experiments.

So people scattered, making a bad situation worse. The outbreak, the biggest recorded in Ebola history, has so far killed more than 670 people in West Africa and is thought to have infected about 400 more, and it shows no sign of abating (see 'Major outbreaks').

Doctors have no cure to offer the infected. Understaffed clinics must make do with isolating infected people, finding and quarantining their families, and educating the public on how to avoid spreading the disease. Although several vaccines and treatments for Ebola do exist, they are stalled in various stages of testing owing to

a lack of funding and of international demand. Even if they did move forwards, it would be years rather than months before the measures would reach the people in need.

For researchers such as Heinz Feldmann, a virologist at the US National Institute of Allergy and Infectious Disease (NIAID) in Hamilton, Montana, the situation seems like it could have been avoided. In 2005, he published a vaccine platform based on vesicular stomatitis virus (VSV) that has since yielded an Ebola vaccine that is effective in macaques (T. W. Geisbert *et al.* *PLoS Med.* **2**, e183; 2005). But money is not available to take the next step — testing the vaccine's safety in healthy humans, says Feldmann. Compared with malaria or HIV, "Ebola is just not that much of a public-health problem worldwide", he says, and consequently draws little interest from public or private funders.

"What works for Ebola is good old-fashioned public health," says Thomas Frieden, director of the US Centers for Disease Control and Prevention in Atlanta, Georgia. "It would be great to have a vaccine, but it's not



Rivers on the run

As the United States destroys its old dams, species are streaming back into the unfettered rivers.

BY RICHARD A. LOVETT

Just outside the small town of Stabler in Washington, hydrologist Bengt Coffin surveys a mountain river that he helped to revive from a decades-long coma.

Today, the clear waters of Trout Creek run fast and cool between banks covered in young alder trees. But just five years ago, an 8-metre-high concrete wall blocked the river at this site. The dam and the reservoir behind it had tamed the river and made it difficult for endangered steelhead trout (*Oncorhynchus mykiss*) to reach their spawning grounds upstream.

In 2009, Coffin led the US Forest Service effort to remove the dam, and Trout Creek has since regained the look of a young river. Vegetation has covered the scars left by the dam and reservoir, and steelhead and other species have started to rebound.

The revival of Trout Creek is part of a growing trend in the United States. About half of the nation's roughly 85,000 known dams no longer serve their intended purposes, and an increasing number are being

removed. Around 1,150 have gone so far, mostly in the past 20 years, according to a tabulation by the watchdog group American Rivers in Washington DC. In an era when many countries are still building dams, the United States is taking them out. "It used to be a crazy idea. Now it's accepted," says Amy Kober, director of communications for American Rivers.

Most of the demolished structures were lower than 5 metres, but in the past few years, projects in the Pacific Northwest have removed much taller ones. At the top end of the spectrum, the US National Park Service is dismantling the 64-metre-high Glines Canyon Dam, the largest of a pair of big dams on Washington's Elwha River. Many of the larger dams were removed because their operators decided that it was too costly to bring the old

An excavator chips away at Washington's Glines Canyon Dam in 2012.

NATURE.COM
For a video and slideshow of dam removals, see: go.nature.com/m1ldiz

structures in line with modern safety and environmental requirements.

The power companies' actions are boons for fish advocates who seek to restore populations of endangered species in the rivers. The dam-elimination trend has also provided an unanticipated research opportunity, because the projects have used diverse approaches to minimize the damage caused by unleashing huge floods of water and decades of accumulated sediment. Some efforts take a slow path, restoring river flow over months or years. Others use explosives and other engineering techniques to drain reservoirs within hours.

Data are still preliminary, but they suggest that both approaches can bring rapid benefits — not just to fish, but also to the habitat on which they depend. The rivers are rebounding at the sites studied so far, says Amy East, a geomorphologist with the US Geological Survey (USGS) in Santa Cruz, California. "We've seen a lot of resilience."

OUT OF COMMISSION

At Trout Creek, Coffin and his colleagues decided to take the cautious route when removing the ageing Hemlock Dam. Built in 1935, the structure provided power and irrigation for a nearby tree nursery that shut down in 1997. It had a fish ladder to allow animals to bypass the dam and swim upstream, but it was poorly built by modern standards and the number of fish using it had steadily declined.

A bigger concern was the reservoir, which had been steadily filling in with silt. By the time the dam was dismantled, the reservoir had become so shallow that it was possible to wade all the way across, says Coffin, waving a hand at mid-thigh level to show the depth of the water. In the mid-summer sun, temperatures in the water could reach 26°C — too warm for steelhead, he says.

When the Forest Service decided to remove the dam, it was particularly concerned about the mud, sand and gravel that had built up in the reservoir. Coffin and others worried that flooding the river with all that sediment would harm the steelhead below the dam in Trout Creek. "All of our baby fish are down there," Coffin says. "We didn't want to decimate them."

The solution was to divert the river into a big pipe and then hire a fleet of dumper trucks to carry away the exposed sediment. In the process, the workers rediscovered the creek's original channel through the reservoir bottom and reinforced its banks with logs to stop them from eroding.

All those efforts seem to have worked. When water was first allowed to flow back through the old reservoir bottom, it initially ran muddy. But just seven hours later, Coffin's team documented the first steelhead venturing into the new channel above the old dam site. "It was that clear," he says.

Since then, the number of steelhead in the river and its tributaries has more than doubled, says fisheries biologist Patrick Connelly at the Columbia River Research Laboratory in Cook, Washington, although he notes that fish populations are variable enough that it will take several years to know whether the trend will continue.

Returning steelhead are not the only signs of success. Just above the old dam site, Coffin winds his way through patches of alder trees that were planted after the dam was removed, then crosses a rocky beach to the river. The rounded stones range from the size of potatoes to loaves of bread, and make for tricky footing. But Coffin is thrilled to see them because none of these ankle-breakers was here when the dam was first taken out. "All of this washed in," he says.

The cobbles provide nesting spots for the trout and a habitat for the insects that the fish eat. "People pay attention to the big animals," Coffin says, "but the bugs are an important part of the system." Reaching into the water, he plucks out a couple of rocks, turns them over and points out six types of insect clinging to the underside, including caddisfly larvae and a stonefly. "The year after the dam was removed, these wouldn't have been here," he says with satisfaction.

Elsewhere in the Pacific Northwest, teams opted for much more

extreme measures to remove the 14-metre-tall Marmot Dam on Oregon's Sandy River in 2007 and the 38-metre-tall Condit Dam on Washington's White Salmon River in 2011.

The dams, both nearly a century old, were too big to take the same approach as at Trout Creek, where it had cost nearly US\$1 million to cart away 42,000 cubic metres of sediment. Marmot had nearly 20 times more sediment and Condit had double that of Marmot. Because it would be too expensive to dig out that material and carry it away, project managers opted for a more radical approach, colourfully described as "blow and go", in which the dams were removed quickly, says Gordon Grant, a research hydrologist at the Forest Service's Pacific Northwest Research Station in Corvallis, Oregon.

The results were impressive — but very different at the two sites. At Marmot, the sediment contained an equal mixture of sand and gravel. Once exposed to river action, it eroded out relatively quickly but sedately, with about half of it gone within 8 months. Researchers were surprised to find that the fish seemed little affected — the first curious salmon poked its nose back towards the former dam site within a day.

At Condit, the sediment contained a higher proportion of fine-grained material: 35% mud, 60% sand and just 5% gravel. The result was predictable in retrospect, but nobody anticipated it.

When engineers blew open a hole at the bottom of the dam, a jet of black liquid shot out as if from a giant fire hose. Instead of the expected flood of water, what came out was more like a mudflow, as waterlogged sediment from the reservoir slumped into the rapidly dropping water, then blasted downriver in a slurry that was as much as 28% sediment by volume. The reservoir lost its water and much of its sediment load in three hours. "It was almost like a volcanic event," says Jon Major, a geomorphologist at the USGS's Cascades Volcano Observatory in Vancouver, Washington. The 5-kilometre-long stretch of river between the dam and its confluence with the Columbia River temporarily became a muddy wasteland. With this kind of approach, says East, the slug of sediment wipes out everything, but the river can start recovering much sooner.

The National Park Service took a much more conservative approach to removing two large dams on the Elwha River, because the stakes were higher. The upstream portions of the Elwha drain more than 100 kilometres of pristine habitat on the north side of Washington's Olympic National Park. A river that large produces a lot of sediment: an estimated 18 million cubic metres was expected to escape from behind the dams, says Jason Dunham, an aquatic ecologist at the USGS office in Corvallis. That is the equivalent of filling eight typical American-football stadiums. And before the dams cut the number of salmon returning each year to around 10,000, the Elwha supported hundreds of thousands of fish.

Unwilling to risk the blow-and-go approach on both dams, engineers opted for a compromise. They quickly removed the lower, 32-metre-high Elwha Dam, which contained only about one-sixth of the total sediment. But the upstream Glines Canyon Dam, which is twice as big, is coming out in a series of steps that have so far lowered it to a 9-metre stub of its former self. East compares the method to deciding whether to uncover a wound quickly or gradually. The approach on the Elwha, she says, is like "pulling the Band-Aid off slowly, over the course of three years".

The good news in these giant projects is that scientists have not seen any serious harm from the feared releases of sediment. Instead, the rivers have proved unexpectedly efficient at flushing the worst of the mud downstream towards the sea, rather than letting it accumulate in river-choking mudflats. "It was not the big catastrophe people thought," East says.

Emily Stanley, a river ecologist at the University of Wisconsin–Madison who has studied dam removals for more than a decade, agrees that it is hard to think of one that had "catastrophically awful" results. (The one exception, she says, was an event in the 1970s, when the demolition of a dam on the Hudson River allowed sediment containing high levels of

"People pay attention to the big animals, but the bugs are an important part of the system."



A reconstructed section of Trout Creek in Washington runs through the site of a former reservoir in 2010, a year after Hemlock Dam was removed.

toxic chemicals called polychlorinated biphenyls (PCBs) to escape from the reservoir and flow downstream.)

Data on the recent dam removals suggest that the fish are now coming back to the unfettered rivers. At Condit, fish were seen returning within weeks of the explosion. Two years later, the total exceeded 5,500, including steelhead and spring Chinook (*Oncorhynchus tshawytscha*), which had been effectively extirpated from the river, says Jody Lando, a quantitative ecologist with Stillwater Sciences in Portland, Oregon, who reported her results in May at an aquatic-sciences meeting in Portland.

Even on the Elwha, where the Glines Canyon Dam still impedes the river, East says that hundreds of salmon have been seen spawning in the lower dam's former lake bed. "That hasn't happened in over a hundred years," she says.

In part, these successes may reflect the fact that the Pacific Northwest is a landscape built by geological disturbances — volcanic outbursts, landslides and floods. Local wildlife has had to adapt to such upheavals, and salmon do that by not always returning to the precise stream of their birth. "There's a fair amount that stray," says East. It is those strays that repopulate any previously inaccessible habitat.

But other parts of the United States have also seen dramatic fish returns. On south-central Wisconsin's Baraboo River, the removal of a string of dams has allowed sturgeon to reach their former spawning grounds. And in New England, the destruction of two dams 7–9 metres high on Maine's Kennebec River and one of its tributaries has allowed Atlantic alewives (*Alosa pseudoharengus*) to repopulate 100 kilometres of previously blocked-off river. In 1999, before the first dam was taken out, no alewives were recorded in the upper part of the watershed, says Serena McClain, head of river restoration for American Rivers. By 2013, the annual run had rebounded to around 3 million.

QUAKE CONCERNS

The next big structure destined for retirement is the 32-metre-tall San Clemente Dam on California's Carmel River. The 93-year-old dam, which was originally built to provide drinking water, is coming out because of concerns over its safety during an earthquake. And there are expensive homes that could be flooded if even modest amounts of sediment were to escape and raise the stream bed, so the dam-removal plan seeks to

avoid that, says East. Instead, the \$84-million project will cut a notch in a ridge near the upstream end of the reservoir, then divert the water into a nearby drainage that rejoins the original river downstream of the dam. "It's a major engineering feat," she says.

Researchers say that the surge in large dam removals in the past ten years has offered valuable insight into how rivers and their ecosystems respond to letting the water flow freely. But because every river and dam is different, it is hard to draw simple lessons that will apply in all situations, says Jim Pizzuto, a fluvial geomorphologist at the University of Delaware in Newark.

Still, the projects have shown that fish are remarkably adept at finding their way back. "If you un-build it, it seems like they will come back," says Grant.

At least, that is the sense emerging from the limited data so far. Researchers are struggling to get detailed statistics on fish recovery — partly because removal projects tend to be planned according to engineering standards, not ones focused on fish and other river residents. And when fish assessments are done, they tend to be carried out by various state and federal agencies that share data only to a limited degree. "A lot of studies wind up on someone's computer, somewhere," says McClain.

But that may be changing because ecological considerations are increasingly part of dam-removal projects. A case in point is Maine's Penobscot River, where a \$62-million public-private partnership is buying dams and removing them to provide better access for fish to more than 1,600 kilometres of the river and its tributaries.

For a country once so bent on taming rivers, attitudes are quickly evolving. At the site of the former Condit Dam, a couple pulls into the car park and walks to a spot overlooking the water. "I come from a dam-building family," says the man. "My father used to build things like this down in California — the Feather River, the Rubicon, the Yuba. I helped."

He pauses.

"A hundred years is a great thing, isn't it? Now we're busily employing people to undo what our ancestors screwed up." He stares silently for a moment at the ribbon of river, flecked with foam, 40 metres below. "It's a great thing." ■

Richard A. Lovett is a freelance writer in Portland, Oregon.

WRITTEN IN BLOOD

DNA circulating in the bloodstream could guide cancer treatment — if researchers can work out how best to use it.

BY ED YONG

In 2012, Charles Swanton was forced to confront one of cancer's dirtiest tricks. When he and his team at the Cancer Research UK London Research Institute sequenced DNA from a handful of kidney tumours, they expected to find a lot of different mutations, but the breadth of genetic diversity within even a single tumour shocked them. Cells from one end differed from those at the other and only one-third of the mutations were shared throughout the whole mass. Secondary tumours that had spread and taken root elsewhere in the patients' bodies were different again¹.

The results confirmed that the standard prognostic procedure for cancer, the tissue biopsy, is woefully inadequate — like trying to gauge a nation's behaviour by surveying a single street. A biopsy could miss mutations just centimetres away that might radically change a person's chances for survival. And although biopsies can provide data about specific mutations that might make a tumour vulnerable to targeted therapies, that information is static and bound to become inaccurate as the cancer evolves.

Swanton and his team laid bare a diversity that seemed insurmountable. "I am still quite depressed about it, if I'm honest," he says. "And if we had higher-resolution assays, the complexity would be far worse."

But researchers have found ways to get a richer view of a patient's cancer, and even track it over time. When cancer cells rupture

and die, they release their contents, including circulating tumour DNA (ctDNA): genome fragments that float freely through the bloodstream. Debris from normal cells is normally mopped up and destroyed by 'cleaning cells' such as macrophages, but tumours are so large and their cells multiply so quickly that the cleaners cannot cope completely.

By developing and refining techniques for measuring and sequencing tumour DNA in the bloodstream, scientists are turning vials of blood into 'liquid biopsies' — portraits of a cancer that are much more comprehensive than the keyhole peeps that conventional biopsies provide. Taken over time, such blood samples would show clinicians whether treatments are working and whether tumours are evolving resistance.

As ever, there are caveats. Levels of ctDNA vary a lot from person to person and can be hard to detect, especially for small tumours in their early stages. And most studies so far have dealt with only handfuls or dozens of patients, with just a few types of cancer. Although the results are promising, they must be validated in larger studies before it will be clear whether ctDNA truly offers an accurate view — and, more importantly, whether it can save or improve lives. "Just monitoring your tumour isn't good enough," says Luis Diaz, an oncologist at Johns Hopkins University in Baltimore, Maryland. "The challenge that we face is finding true utility."

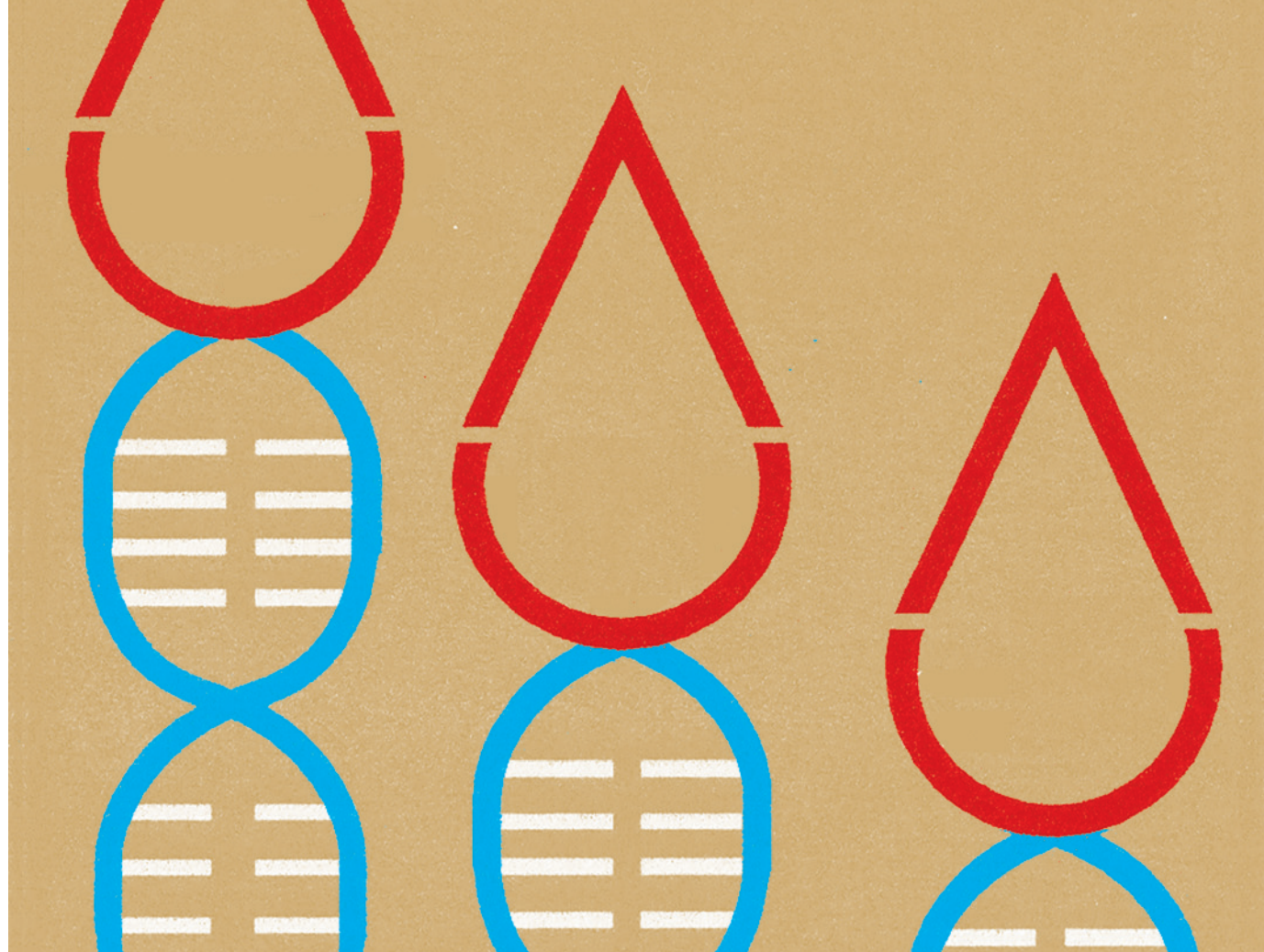
If researchers can clear those hurdles, liquid biopsies could help clinicians to make better choices for treatment and to adjust those decisions as conditions change, says Victor Velculescu, a genetic oncologist at Johns Hopkins. Moreover, the work might provide new therapeutic targets. "It will help bring personalized medicine to reality," says Velculescu. "It's a game-changer."

DELAYED ACTION

Scientists first reported finding DNA circulating in human blood in 1948 (ref. 2), and specifically in the blood of people with cancer in 1977 (ref. 3). It took another 17 years to show that this DNA bore mutations that are hallmarks of cancer — proof that it originated from the tumours^{4,5}.

The first practical use of circulating DNA came in another field. Dennis Lo, a chemical pathologist now at the Chinese University of Hong Kong, reasoned that if tumours could flood the blood with DNA, surely fetuses could, too. In 1997, he successfully showed that pregnant women carrying male babies had fetal Y chromosomes in their blood⁶. That discovery allowed doctors to check a baby's sex early in gestation without disturbing the fetus, and ultimately to screen for developmental disorders such as Down's syndrome without resorting to invasive testing. It has revolutionized the field of prenatal diagnostics (see *Nature* 507, 19; 2014).

"Cancer has been slower to catch on," says



Nitzan Rosenfeld, a genomicist at the Cancer Research UK Cambridge Institute. This is partly because tumour DNA is much harder to detect than fetal DNA. There is typically less of it in the blood, and the amounts are extremely variable. In people with very advanced cancers, tumours might be the source of most of the circulating DNA in the blood, but more commonly, ctDNA makes up barely 1% of the total and possibly as little as 0.01%. Early sequencing technologies were not up to the task of detecting it — at least, not consistently or reliably enough to use ctDNA as a biomarker.

But the past decade has brought sensitive techniques that can detect and quantify minute amounts of DNA. For example, an amplification method known as BEAMing — which fastens circulating DNA to magnetic beads that can then be isolated and counted — can detect ctDNA even if it is outnumbered by healthy cell DNA by a factor of 10,000 to 1.

Genetic oncologists Bert Vogelstein and Kenneth Kinzler at Johns Hopkins developed the technique, and in 2007 they described⁷ using it to track ctDNA in 18 people who were being treated for bowel cancer. After surgery, the patients' ctDNA levels fell by 99%, but in many cases the signal did not disappear completely. In all but one of the people with detectable ctDNA at the first follow-up appointment, the tumours eventually returned. None of the people with undetectable levels after surgery experienced a recurrence.

These results suggested that ctDNA can reveal how well a patient has responded to surgery and whether they need chemotherapy to finish off any lingering cancer cells. Researchers soon found similar results for other types of cancer. Rosenfeld and his Cancer Research UK colleagues James Brenton and Carlos Caldas showed that ctDNA provides a precise portrait of advanced ovarian and breast cancers⁸. And in the largest study yet, Diaz and other members of the Johns Hopkins group detected ctDNA in at least 75% of patients with advanced tumours, in organs as diverse as the pancreas, bladder, skin, stomach, oesophagus, liver and head and neck⁹. (Brain cancers were a notable exception, because the blood–brain barrier stops tumour DNA from reaching the bloodstream.)

BETTER BIOMARKERS

Circulating DNA might perform better than the protein biomarkers that researchers have been seeking and refining for decades. Proteins are used in the clinic to diagnose illnesses and monitor people undergoing treatment. For example, prostate-specific antigen is a biomarker for prostate cancer, but it can give false positives because there are other reasons that the antigen can be elevated in the blood. False positives should be rarer with ctDNA because it is defined by mutations and other genomic changes that are hallmarks of cancer cells. And although most protein biomarkers stay in the blood for weeks, ctDNA has a half-life of less

than two hours, so it gives a clearer view of a tumour's present, rather than its past. The Cambridge and Johns Hopkins teams have found that ctDNA is more sensitive than protein biomarkers when it comes to detecting breast¹⁰ and bowel⁹ cancers, respectively, and it is more accurate at tracking tumour disappearance, spread and recurrence.

Both teams also showed that ctDNA was more sensitive than circulating tumour cells — intact cancer cells that also travel around the bloodstream and have been an intense area of research. In a sub-study of 16 people, Diaz's team found that where both were present, ctDNA fragments outnumbered circulating tumour cells by 50 to 1 (ref. 9). And although ctDNA was always there if the circulating cells were, 13 people with detectable tumour DNA had no trace of such cells.

But most exciting to scientists, says Diaz, is the ability to watch tumours evolve and adapt over time: "It'll help us answer questions in oncology that have never been answered before."

For example, why do so many targeted therapies eventually fail? Gefitinib and panitumumab are among several drugs that block the epidermal growth factor receptor (EGFR), a protein involved in cell growth and division that is overactive in a number of cancers. People taking these drugs do very well — briefly. But after a few months, their cancers almost always develop resistance, often through

changes to other genes, such as *KRAS*, which is mutated in many cancers.

To monitor patients and decide on the next course of action, clinicians would normally need to take multiple biopsies. But people with advanced cancer often have several tumours to test, and different parts of any single tumour could be resistant in different ways. Biopsies are invasive and risky, and difficult for inaccessible and fragile organs such as the lungs. “You can’t just go to the patient and get five more biopsies after the treatment fails,” says Velculescu. Taking blood is simple in comparison.

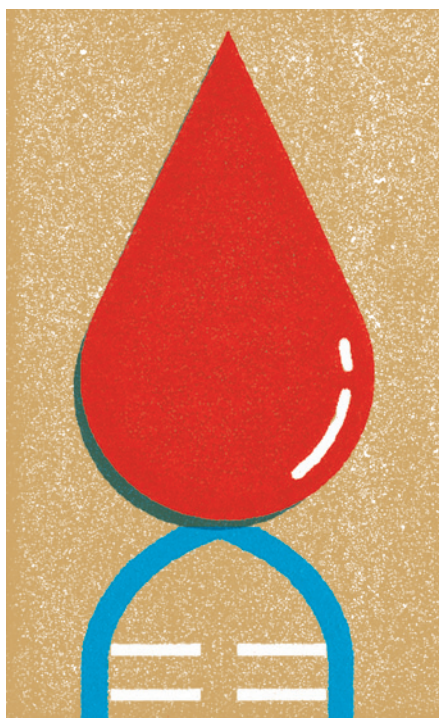
In 2012, Diaz’s team reported¹¹ using ctDNA to study patients who were being treated with EGFR inhibitors. The researchers found 42 different *KRAS* mutations that confer resistance; on average, these turned up 5 months before imaging techniques showed that the tumours were progressing. The team was specifically looking for *KRAS* mutations, but Rosenfeld’s group has used ctDNA to identify resistance mutations from a blind start. Last year, the researchers described how they had sequenced the complete exomes — the 1% of the genome that encodes protein — in blood samples from six people being treated for advanced breast, lung or ovarian cancers. In five cases, the unguided search revealed routes to resistance, such as mutations that prevent drugs from binding to their target proteins¹².

Spotting resistance early would let clinicians take patients off toxic and expensive drugs that are unlikely to keep working. And by identifying the mutations that underlie the resistance, they could find effective alternatives or drug combinations. “The hope is that we can turn cancer from a deadly disease into a chronic one,” says Velculescu. “You treat someone with one therapy and when it stops working, you switch, or alternate back and forth.”

CLINICAL CAVEATS

Despite its promise, ctDNA is not yet ready for a starring role in the clinic. For one thing, the most sensitive techniques for detecting it, such as BEAMing, rely on some knowledge of which mutations to look for. This knowledge can be provided by taking a biopsy, sequencing its mutations, designing patient-specific molecular probes that target them, and using those probes to analyse later blood samples — a laborious approach that must be repeated for each patient. The alternative is to use exome sequencing, as Rosenfeld’s team did. This requires no previous knowledge about the cancer, but it is prohibitively expensive to sequence and analyse every sample at the depth required to detect rare mutant fragments.

Maximilian Diehn, a radiation oncologist at Stanford University in California, has tried to combine the best of both worlds. His team identified a small proportion of the genome — just 0.004% — that is repeatedly mutated in lung cancers¹³. Whenever the researchers get a new blood sample, they sequence this fraction



**“IT’LL HELP US
ANSWER QUESTIONS IN
ONCOLOGY
THAT HAVE NEVER
BEEN ANSWERED
BEFORE.”**

10,000 times over. This picks up even rare mutant fragments, and the focused approach keeps costs down. Because almost everyone with lung cancer has at least one mutation in these regions, the method should work in almost every patient, says Diehn. The team is now working to develop similar mutation panels for other types of cancer, and to validate the technique in clinical trials — work that could take several years.

Like practically all ctDNA biopsy techniques, Diehn’s approach does not do well at picking up early forms of cancer. In a small study¹³, it detected every lung cancer of stage II or higher, but only half of stage I tumours. This is understandable — advanced cancers simply discharge more DNA — but it limits ctDNA’s potential as a cancer-screening tool.

Diehn says that more-sensitive techniques could overcome this problem, but Diaz disagrees. “The limiting factor is biology,” he says. “There just aren’t a lot of fragments

in circulation.” And if ctDNA hints at the presence of an undetected cancer, what then? “If you detect a mutation in the circulation, you don’t know where it’s coming from,” says Diaz.

There are other unknowns, too. Does ctDNA paint a truly representative portrait of a cancer? Do tumours that have spread to other organs release as much DNA as the original tumours? Do all the cells in a tumour release as much ctDNA as each other? Diaz says that the only way to answer these questions is to do ‘warm autopsies’ — to take samples and characterize all of a person’s tumours very soon after death, and compare them with ctDNA extracted in life. “This is the heavy lifting that’ll need to be done in the field,” he says.

And the biggest question remains: does an accurate picture of tumour burden, or a real-time look at emerging mutations, actually save patients or improve their quality of life? Even if doctors discover that someone’s tumour has developed a resistance mutation, that insight is useless if there are no drugs that target the mutation. “The limitation is the reality of targeted therapies,” says Velculescu. “You get all this information — but so what? Our approaches to understanding cancer are outstripping our clinical options.”

Even if ctDNA does not yet affect outcomes, scientists say that it is an invaluable research tool, and clinicians are starting to collect it routinely. Swanton, for example, is leading a £14-million (US\$24-million) lung-cancer study called TRACERx (Tracking Cancer Evolution Through Therapy), which will use both conventional biopsies and ctDNA collected once every three months. The circulating DNA may or may not provide clues that help the study participants, but at the very least, it will give Swanton a much better understanding of how lung cancer evolves, and how to control that evolution.

As Rosenfeld argues, it is better to have this information than not to. Currently, he says, “we’re groping in the dark. Why would you do that if you have a tool that allows you to see what’s happening?” ■

Ed Yong is a science journalist based in London.

1. Gerlinger, M. *et al.* *N. Engl. J. Med.* **366**, 883–892 (2012).
2. Mandel, P. & Metais, P. C. R. *Séances Soc. Biol. Fil.* **142**, 241–243 (1948).
3. Leon, S. A., Shapiro, B., Sklaroff, D. M. & Yaros, M. J. *Cancer Res.* **37**, 646–650 (1977).
4. Vasioukhin, V. *et al.* *Br. J. Haematol.* **86**, 774–779 (1994).
5. Sorenson, G. D. *et al.* *Cancer Epidemiol. Biomarkers Prev.* **3**, 67–71 (1994).
6. Lo, Y. M. D. *et al.* *Lancet* **350**, 485–487 (1997).
7. Diehl, F. *et al.* *Nature Med.* **14**, 985–990 (2008).
8. Forshaw, T. *et al.* *Sci. Transl. Med.* **4**, 136ra68 (2012).
9. Bettegowda, C. *et al.* *Sci. Transl. Med.* **6**, 224ra24 (2014).
10. Dawson, S.-J. *et al.* *N. Engl. J. Med.* **368**, 1199–1209 (2013).
11. Diaz, L. A. Jr *et al.* *Nature* **486**, 537–540 (2012).
12. Murtaza, M. *et al.* *Nature* **497**, 108–112 (2013).
13. Newman, A. M. *et al.* *Nature Med.* **20**, 548–554 (2014).

COMMENT

BEHAVIOUR Energy research needs social science **p.529**



NEUROSCIENCE Has the power of mirror neurons been oversold? **p.532**

METRICS Total grant income is a questionable proxy for research quality **p.534**

CONSERVATION Don't be too hasty to wipe out invasive toads in Madagascar **p.534**

JIANAN YU/REUTERS/CORBIS



A girl drinks water from a tap in Anhui province, China.

A sustainable plan for China's drinking water

Tackling pollution and using different grades of water for different tasks is more efficient than making all water potable, say **Tao Tao** and **Kunlun Xin**.

Making drinking water safe is a priority in China. Serious health and social problems concentrate in areas where the water quality is poor. Every year, 190 million people in China fall ill and 60,000 people die from diseases caused by water pollution such as liver and gastric cancers¹. Around 300 million people face shortages of drinking water². In a 2009 nationwide assessment, one-quarter of 4,000 urban water-treatment plants surveyed did not comply with quality controls, stoking public fears about the health impacts.

The Chinese government is in the

middle of a five-year 410-billion-renminbi (US\$66-billion) programme to deliver safe drinking water to all town and city residents — about 54% of the population — by 2015. The focus is on upgrading 92,300 kilometres of mains pipes and thousands of water-treatment plants to developed-world standards.

But this infrastructure-focused approach is ill-suited to China, which is projected to remain a developing country until at least 2050. Urban expansion will outpace improvements to public water systems, and treating polluted water will require large

amounts of energy, expensive technologies and chemicals.

Instead, the government should focus on cleaning water sources and recycling water. The first priority must be to purge rivers and lakes of industrial and agricultural pollutants, and to prevent these from entering the water table in the first place. Cheaper technologies at the point of use, such as purifiers on taps, would be enough to deliver clean drinking water to most of China's population, because drinking water accounts for only a few percent of total consumption (see 'China's water'). Water of lower quality can suffice ►

► for laundry, bathing and kitchen use.

In 2012, China mandated that tap water in all cities should meet a standard based on 106 indices called for by the World Health Organization. As well as investing in water treatment, distribution and quality monitoring, the government has added drinking-water safety to its list of 13 major technology projects, which include lunar-exploration and crewed space programmes. Billions have been spent on researching drinking-water problems in key river basins and lakes. But so far, only a few cities meet the desired standard.

SHORT SUPPLY

It is a problem of supply and demand. Supply is a challenge because almost half of China's water sources are polluted. Wells and aquifers are contaminated with fertilizers and pesticide residues and heavy metals such as arsenic and manganese from mining, the petrochemical industry and domestic and industrial waste. More than three-quarters (76.8%) of 800 wells monitored in nine provinces, plus autonomous regions and municipalities, including Beijing, Shanghai and Guangzhou, failed to meet standards for groundwater in a 2011 national evaluation³.

Water demand is a challenge because of runaway economic growth and urbanization. China is short of 40 billion tonnes of water a year on average. In 2011, 665 cities consumed in excess of 44 billion tonnes of water, or an average of around 66 million tonnes each. By 2020, when China's urban population proportion is projected to reach 60%, cities might need 58 billion tonnes of water.

But take a closer look at how that water is used, and the problem becomes tractable. Almost two-thirds of municipal water is used by industry, agriculture and construction. Households consume the remaining third (365 million people used 15.3 billion tonnes of water in 2011). Of that,

laundry, bathing and washing up take up most (together more than 80%). Cooking and drinking use just over 2% (1.1 billion tonnes). In other words, most household water need not be fit to drink.

Bringing a large developing country such as China up to the same standard as a developed country will require more-intensive water treatment. This has environmental consequences. In Jiangsu province, for example, carbon dioxide emissions increased by 28% in 2012 when a type of water filtration called ozone-biological activated carbon treatment was extended to one-quarter of the province's supply

(5.3 million tonnes per day). China is in need of cheap, energy-efficient methods of water purification that minimize chemical use⁴.

Even if tap water becomes drinkable, few people will stop boiling drinking water, a habit that is ubiquitous in China. Boiling kills or deactivates all waterborne pathogens, including protozoan cysts such as cryptosporidium that can be resistant to chemical disinfection, and viruses such as rotavirus and norovirus that are too small to filter out⁵. Even if the water is turbid, boiling can remove microorganisms and volatile organic compounds such as benzene and chloroform.

In this context, purification systems that improve drinking water at the point of use are a good fit. In Kenya, Bolivia and Zambia, water purifiers have been shown^{6,7} to reduce diarrhoeal disease by 30–40%. Fewer than 5% of Chinese homes currently have these, despite a unit costing only around 1,500 to 2,000 renminbi.

China's water-purification industry is growing by about 40% a year — fewer people are buying water dispensers and barrelled water. But water-purification devices are

“China is short of 40 billion tonnes of water a year on average.”

unregulated. Incomplete after-sales service leads to improper maintenance; delays in changing filter cartridges can introduce microorganisms. Filters and units made from toxic materials such as non-food-grade plastic are ineffective.

Treated grey water (waste water from showers and baths) and black water (from toilets) are increasingly used in China for industrial and irrigation purposes, and for flushing in new residences. But this type of recycling is impractical for most existing households, owing to the high cost and disruption of installing the necessary plumbing.

SOME FOR ALL

What next? By using cheap, low-carbon water purifiers in all homes, China can avoid the technology 'lock-in' that leads developed countries to waste potable water, and leap-frog to a sustainable supply system. In the long term, the improvement of water sources will ensure that most people have safe drinking water.

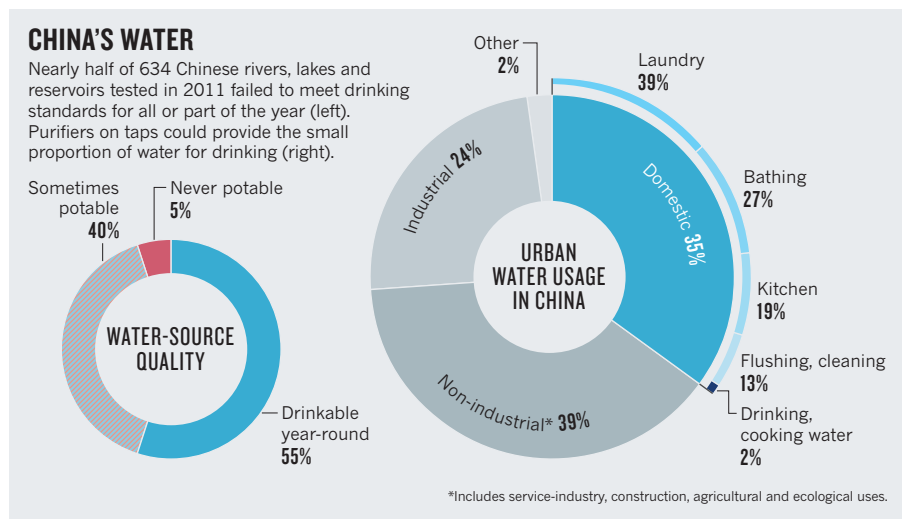
The local and national governance structures overseeing water supply and water-source pollution must be merged and an agency created to manage them — currently these are administered separately, by the Ministry of Housing and Urban-Rural Development, Ministry of Water Resources and the Ministry of Environmental Protection. Responsibilities need to be clearly defined.

Regulation and standards for water purifiers should be enforced. Business models for scaling up these technologies require investigation and testing of factors including the costs of production and implementation, pricing, subsidy and microfinance⁷.

As the Global Consultation on Safe Water and Sanitation for the 1990s⁸ put it, China's aim in supplying clean water should be: “Some for all rather than more for some”. ■

Tao Tao is professor and **Kunlun Xin** is associate professor at the College of Environmental Science and Engineering, and the Key Laboratory of Yangtze River Water Environment Tongji University, Shanghai, China.
e-mails: taotao@tongji.edu.cn; xkl@tongji.edu.cn

1. Qiu, J. *Science* **334**, 745 (2011).
2. Qiu, J. *Nature* <http://dx.doi.org/10.1038/news.2009.111> (2009).
3. Ministry of Water Resources of the People's Republic of China *China Water Resources Bulletin 2011* (China Water & Power Press, 2012).
4. Shannon, M. A. et al. *Nature* **452**, 301–310 (2008).
5. Clasen, T. F., Thao, D. H., Boisson, S. & Shipin, O. *Environ. Sci. Technol.* **42**, 4255–4260 (2008).
6. Fewtrell, L. et al. *Lancet Infect. Dis.* **5**, 42–52 (2005).
7. Sobsey, M. D. et al. *Environ. Sci. Technol.* **42**, 4261–4267 (2008).
8. Gadgil, A. *Annu. Rev. Energy Environ.* **23**, 253–286 (1998).





Masai women from Kenya take a course on solar energy in India.

Energy studies need social science

A broader pool of expertise is needed to understand how human behaviour affects energy demand and the uptake of technologies, says **Benjamin K. Sovacool**.

To secure a safe, reliable and low-carbon energy future, we must alter both technologies and human behaviour¹. The US Department of Energy notes² that supply and demand is “affected as much by individual choice, preference, and behavior, as by technical performance”.

Yet many researchers and policy-makers continue to focus on only one side of the energy dilemma. In the United States, for every dollar in research funds spent on behavioural and demand-side energy research, \$35 is spent on energy supply and infrastructure³. Social sciences, humanities, and the arts are marginalized in energy research, and major statistical agencies do not usually collect qualitative data about energy consumption. Similar problems are apparent in Europe⁴.

My analysis of the peer-reviewed energy-research literature shows how biases handicap the field⁵. Engineers and economists are ignoring people and miscasting decision-making and action. Academic researchers

frequently obsess over technical fixes rather than ways to alter lifestyles and social norms⁶. Interdisciplinary research remains stymied by institutional barriers in academia and government⁷. National and local energy bodies have conventionally had few social scientists on staff⁸. And most leading journals in the field focus on one discipline.

Now the energy field needs to learn from health, agriculture and business, and bring together social and physical scientists. Universities should develop courses focused on solving energy problems, granting agencies should prioritize and direct more money to behavioural work, and energy journals should broaden their scope. Already, there are promising examples of how inclusive and interdisciplinary energy research can encourage energy efficiency, and so address global environmental challenges such as climate change⁹.

I examined the authorship and scope of 4,444 full-length articles over 15 years (1999 to 2013) in three leading energy technology

and policy journals: *Energy Policy* and *The Energy Journal* have high impact factors, and *The Electricity Journal* was included to sample a regulatory journal. I found four worrisome trends: an undervaluation of the influence of social dimensions on energy use; a bias towards science, engineering and economics over other social sciences and the humanities; a lack of interdisciplinary collaboration; and the under-representation of female authors or those from minority groups.

For instance, technology adoption, the complexity of choice-making, and the human dimensions of energy use and environmental change were rarely covered (see ‘Neglected topics’). Most articles (85%) focused on advanced energy-production systems, such as nuclear reactors, sources of renewable electricity and biofuels, or the technical elements of electricity generation, transmission and distribution — hardware — rather than the human ‘software’ behind it. Simple devices such as cooking stoves, bicycles, light bulbs and distributed generation were studied in less than 3.5% of articles. Behaviour and energy demand was investigated in less than 2.2% of papers. If this work is being published, it is in environmental sociology, psychology and political-science journals that few energy researchers read.

SOCIAL OUTCASTS

Social-science authorship and citations are also relatively low (see ‘Publishing trends’). Science, engineering, economics and statistics account for more than half (67%) of institutional affiliations as reported by authors; non-economic social science for less than 20%. Sociology, geography, history, psychology, communication studies and philosophy each constituted less than 0.3% of author affiliations.

References to social-science and humanities journals, with their insights into how consumers and politicians behave, were less than 4.3% of 90,097 citations across the sample. Little research took place in the ‘real world’. Most studies are the result of work undertaken at the bench or desk using computer models and experiments, rather than field research, interviews and surveys.

Another trend is that the scientists and engineers writing in these journals rarely collaborate beyond their fields. About half of published authors in the sample wrote alone and one-quarter published with colleagues within their discipline. Less than 23% of articles involved interdisciplinary collaborations between authors.

Furthermore, the vast majority of authors hail from affluent Western institutions and countries where research money is abundant. They focus on problems facing the industrialized world. Of the 9,549 authors who listed their country of residence, 87%

came from either North America or western Europe. African, Asian, Latin American and Middle Eastern authors were few. Authors were mostly male: only 15.7% could be identified as female. Norms of authorship and collaboration vary, but these trends held for each year examined: female authorship remained below 17.4% and non-Western authorship under 16%, for example.

FIVE RECOMMENDATIONS

To bring in social scientists and other marginalized researchers, I have five recommendations.

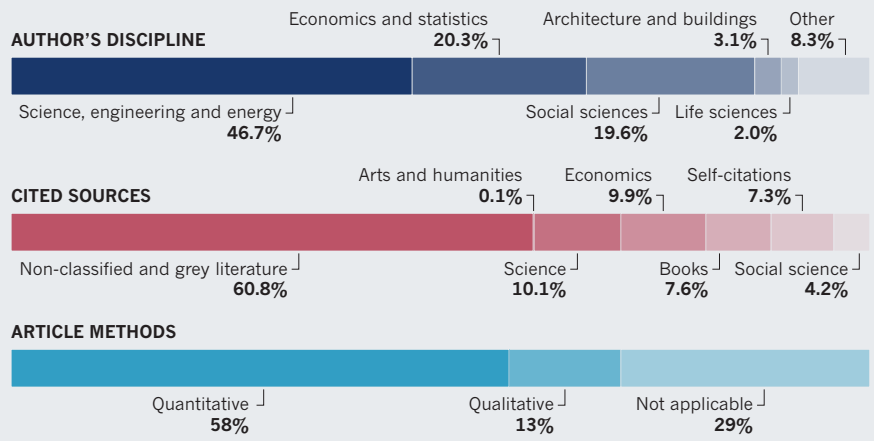
First, public and private organizations should overhaul the way they structure and disburse funding for energy research and development. They should give a bigger slice to social scientists, improve incentives for interdisciplinary work and prioritize social topics in their funding calls — such as the perceptions of energy users, the needs of people affected by energy production and prevailing customs, traditions and behaviours.

Second, to reduce disciplinary bias, energy ministries, statistical agencies and public utility commissions should focus more on energy behaviour and demand, rather than just supply. Delaware and the District of Columbia, for instance, have sustainable-energy utilities, which advise residents about behavioural changes they can make to save energy and money. The statewide energy-efficiency utility, Efficiency Vermont, provides funding and behavioural guidance to homes, farms and factories.

Third, administrators should make energy research more problem-oriented, including social perspectives as a matter of course. Universities should develop topical programmes on energy, as they have in agricultural research, medicine and business. Curricula might include efficient and sustainable consumption, risk management, public decision-making and the design of technologies for public acceptance and

PUBLISHING TRENDS

Social-science studies were rarely published in three leading energy journals from 1999 to 2013. The emphasis on technology rather than human behaviour in energy research is reflected in the disciplinary backgrounds of authors, work referenced, and methods used.



use. Good examples include the University of Edinburgh, UK, which offers an interdisciplinary master's degree in climate accounting; Aarhus University in Denmark has a business-development degree that combines engineering, innovation studies, energy studies, business and marketing; and Carnegie Mellon University in Pittsburgh, Pennsylvania, has an engineering and public policy department. Outside academia, the US Defense Advanced Research Projects Agency has successfully used a 'challenges-centred' approach to national-security problems since it was created in 1958.

Fourth, researchers should do more to accommodate expertise and data from laypersons, indigenous groups, community leaders and other non-conventional participants. Although this may require special training to do effectively, such interactions would encourage greater feedback and integrate diverse viewpoints.

Fifth, journal editors can prioritize interdisciplinary, inclusive, comparative,

mixed-methods research. A new journal published by Elsevier, *Energy Research & Social Science* (of which I am editor-in-chief), calls explicitly in its aims and scope for papers that blend disciplinary concepts, go beyond single case studies, and utilize an assortment of methods. *Wiley Interdisciplinary Reviews: Energy and Environment* also seeks cross-disciplinary assessments of energy systems.

Energy studies must become more socially oriented, interdisciplinary and heterogeneous. Problem-focused research activities that centre on both physical and social processes, include diverse actors and mix qualitative and quantitative methods, have a better chance of achieving analytic excellence and social impact. ■

Benjamin K. Sovacool is professor of business and social sciences, and director of the Centre for Energy Technologies, at Aarhus University in Denmark. He is also associate professor of law at Vermont Law School in South Royalton, Vermont, USA, and editor-in-chief of *Energy Research & Social Science*.
e-mail: sovacool@vt.edu

1. Dietz, T., Gardner, G.T., Gilligan, J., Stern, P.C. & Vandenbergh, M. P. *Proc. Natl Acad. Sci. USA* **106**, 18452–18456 (2009).
2. US Department of Energy *Report on the First Quadrennial Technology Review* 125 (US Department of Energy, 2011).
3. Gaffigan, M. E. *Advanced Energy Technologies: Budget Trends and Challenges for DOE's Energy R&D Program* (US Government Accountability Office, 2008).
4. Goldblatt, D. L. et al. (eds) *Tackling Long-Term Global Energy Problems: The Contribution of Social Science* (Springer, 2012).
5. Sovacool, B. K. *Energy Res. Soc. Sci.* **1**, 1–29 (2014).
6. D'Agostino, A. L. et al. *Energy* **36**, 508–519 (2011).
7. Lutzenhiser, L. & Shove, E. *Energy Policy* **27**, 217–227 (1999).
8. Stern, P. C. *Science* **260**, 1897–1899 (1993).
9. Alcott, H. & Mullainathan, S. *Science* **327**, 1204–1205 (2010).

NEGLECTED TOPICS

Twelve subjects seldom considered in energy studies.

Topic	Example
Gender and identity	Pollution from cooking stoves posing greater risk to women than men
Philosophy and ethics	Future generations bearing the burden of pollution
Communication and persuasion	Energy information changing individual or firm behaviour
Geography and scale	Mismatching the size of energy systems to patterns of demand
Social psychology and behaviour	Shaping energy choices by trust, control and denial
Anthropology and culture	Temporal and regional differences in conceptions of energy services
Research and innovation	How people, markets and institutions drive innovation
Politics and political economy	Resources contributing to conflict or stymying growth
Institutions and energy governance	Evolving rules and norms to address collective energy problems
Energy and development	Energy use contributing to economic growth and falling poverty
Externalities and pollution	Costs to society of erosions of environmental and ecological capital
Sociology of technology	Economic, political and social drivers of energy consumption



It is not necessary to mimic another's actions to understand them.

NEUROSCIENCE

Looking-glass wars

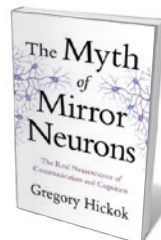
Patricia Smith Churchland welcomes a critique of the mirror-neuron theory linking brain and behaviour.

Well-grounded theories that connect neurons with behaviour are highly prized but in short supply. One behaviour we would dearly love to explain is how humans, along with some birds and mammals, 'mind-read' — that is, attribute mental states such as goals, intentions and feelings to others, to predict and understand their actions. Because others' motivations are not directly observable, the capacity to intuit them has seemed to require some special explanation.

The discovery several decades ago of 'mirror' neurons in the premotor cortex of macaque monkeys spawned the idea that these neurons provide that special explanation. As described by neuroscientists Giacomo Rizzolatti and Laila Craighero, mirror neurons respond both when one monkey sees another make a certain movement and when the animals make the same movements themselves.

The idea that a brain's capacity to mind-read emerges more or less automatically from

the activity of mirror neurons was articulated thus by neuroscientist Marco Iacoboni in *Mirroring People* (Farrar, Straus and Giroux, 2008): "Mirror neurons undoubtedly provide, for the first time in history, a plausible neurophysiological explanation for complex forms of social cognition and interaction." Examples of such social cognition might be my knowing what you intend when I see you head for the chicken coop with an axe in hand, or my understanding what a baseball player feels when he strikes out in the last inning. In *The Myth of Mirror Neurons*, cognitive scientist Gregory Hickok undertakes a balanced and detailed examination of claims that have flourished in the past ten years — that



The Myth of Mirror Neurons: The Real Neuroscience of Communication and Cognition
GREGORY HICKOK
W. W. Norton: 2014.

mirror neurons are the key to explaining our capacity for reading other minds.

The mirror-neuron approach to mind-reading depends on the assumption that our own inner lives are transparently revealed to our own minds. So when we see someone holding an axe while heading to the hen house, that observation activates in us not only the motor programme for that movement, but also the mental state that is its normal antecedent — intent to slaughter a chicken. Simulation of observed behaviour, so this argument goes, allows us to identify others' intentions. Iacoboni declares that mirror neurons "are at the heart of how we navigate through our lives".

These are bold and promising ideas, and Hickok wants to know whether the research makes good on the promise. One basic problem that he sees is this: the evidence shows that mirror neurons respond to movements, one's own and others'. The claim is that mirror neurons reflect high-level understanding of goals. But how? Rizzolatti and his colleagues tried to address this matter by designing an experiment in which a monkey is trained either to put food in its mouth or to put an object in a cup stationed near its mouth: similar movements, different goals.

When other monkeys witness these actions, the responses of mirror neurons in their inferior parietal lobes vary depending on whether the action-performer grasps to eat or to place. So are these neurons sensitive to observing similar movements with different goals? Here things get complicated because the non-food object was taken from a jar, but the food was not. In Hickok's view, this leaves open the question of whether the mirror neurons are reading the goal or merely responding to different movements.

As Hickok sizes it up, the responses of the witness monkeys' mirror neurons seem to be explicable in terms of past associations, implying that the claim of mind-reading through simulation is superfluous. Likewise with my expectations regarding the unfortunate hen: my brain does not need to produce a simulation of your behaviour to know your motive, because in the past I have detected axe-wielding in the vicinity of chickens before they are slaughtered. Adding to the scepticism, Hickok points out that you can understand many actions that you never perform. Hickok's dog, who never throws the ball himself and thus cannot simulate ball-throwing, nevertheless reliably predicts the ball's trajectory by watching Hickok's arm. As for simulating feelings, I may know that a baseball player is disappointed after striking out, but feel only joy if it means my team is winning.

Some scientists sought to draw support for the simulation hypothesis from the motor theory of speech perception (MTSP). In brief, the idea of MTSP, popular in the 1950s, is that I can understand what you

SILVIA OTTE/GETTY

mean when you say, “The cat is swimming” by recreating that bit of speech in my brain’s speech area. Language is Hickok’s area of expertise, and he reminds us of the experiments that saw MTSP shelved. For example, people with a disorder called Broca’s aphasia are unable to produce speech, but they can still understand it, as can children in the pre-speech language-learning phase of development and people born with cerebral palsy who have severely impaired speech production. Some researchers dismiss those flaws in MTSP on the grounds that the mirror-neuron story explains language understanding. The circularity here is not reassuring.

Not least of the problems with the mirror-neuron approach is that learning mind-reading skills cannot be just a matter of simulation, because such skills depend on a co-evolution of understanding of the self and of others. Recognition of one’s own inner states is not a computational freebie.

How fares the hypothesis that autism is fundamentally a mirror-neuron disorder? So far, it is mixed. A deeper perspective derives from post-mortem studies of the brains of youngsters with autism. These show patches of laminar disorganization — types of neuron in the wrong layer making the wrong connections — in wide swathes of the prefrontal cortex, including areas important for executive function, motor control and social cognition, as well as areas that probably contain some mirror neurons. This suggests that autism is not primarily or essentially a disorder of a hypothetical mirror-neuron system, but a broader disorder that affects many aspects of normal brain function, including cognition.

Hickok does not for a moment deny that we mind-read. Rather, his point is that the roles of mirror neurons and simulation have been oversold. The upshot of his inquiry is an analogue of the familiar warning: if it seems too good to be true, it probably is.

Hickok’s critique deserves to be widely discussed, especially because many scientists have bought into the mirror-neuron theory of action understanding, perhaps because they lack the time or inclination to peer into its workings themselves. Hickok performs a valuable service by laying out the pros and cons clearly and fairly. He ends by agreeing that although mirror neurons may well have a role in explaining communication and empathy, many other neural networks with complex responses are undoubtedly involved. Those networks and their roles are still to be clarified. ■

Patricia Smith Churchland is professor emerita of philosophy at the University of California, San Diego, and an adjunct professor at the Salk Institute in La Jolla, California. She is the author of *Braintrust* and *Touching a Nerve*.
e-mail: pschurchland@ucsd.edu

Books in brief



Invisible: The Dangerous Allure of the Unseen

Philip Ball **BODLEY HEAD (2014)**

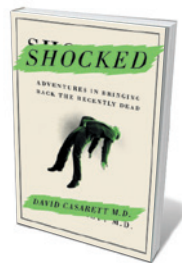
Young children, notes science writer Philip Ball, believe they vanish when they shut their eyes. Such beliefs wither, but “the dream and the desire” for invisibility remain, and Ball traces these through history. The urge has spawned occultism, stage magic, a fascination with camouflage, and legends centring on rings and cloaks. It re-emerged a century ago in the confluence of paranormal beliefs and the new physics — and, today, in optical physicists’ invisibility shields. Ball argues that this “mythical lens” we train on reality inspires scientific discovery, but we need to understand its calibration.



H is for Hawk

Helen Macdonald **JONATHAN CAPE (2014)**

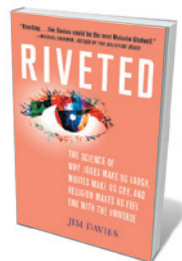
This extraordinary book is ostensibly about falconry. It actually tells how a human wild with grief came to fathom a wild mind — a process in which the question of who was being tamed was always up in the air. Writer Helen Macdonald, devastated by her father’s death, took on a goshawk. Her narrative interweaves exquisitely rendered observations — of hawk behaviour, her immersion in the bird’s world and what happens between them — with the life and work of author T. H. White, whose 1951 *The Goshawk* inspired her as a child. Soars beyond genres, and burns with emotional and intellectual intensity.



Shocked: Adventures in Bringing Back the Recently Dead

David Casarett **CURRENT (2014)**

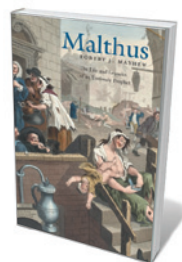
In 1986, a toddler named Michelle Funk drowned and lay dead for three hours before a medical team coaxed her back to life. Decades later, relates physician David Casarett, the science of resuscitation is very much alive. In this disarmingly amusing investigation, Casarett covers breakthroughs, devices, hazards and case studies. He visits resuscitative techniques of the past, such as blowing tobacco smoke into the victim’s rectum; the cellular effects of methods using electricity and low temperature; and potential future advances, including reducing metabolism.



Riveted: The Science of Why Jokes Make Us Laugh, Movies Make Us Cry, and Religion Makes Us Feel One with the Universe

Jim Davies **PALGRAVE MACMILLAN (2014)**

Moments that jolt or delight us punctuate our lives. But whereas shock might be salutary in an art gallery, it can trigger blind belief in other contexts, points out cognitive scientist Jim Davies. Expounding his theory of ‘compellingness foundations’, Davies synthesizes research on what makes us susceptible to gripping stimuli, such as our drives to discover patterns and to find incongruity, and our attraction to hope and fear. Scepticism, he argues, can help us to build resistance to riveting ideas that turn out to be duds.



Malthus: The Life and Legacies of an Untimely Prophet

Robert J. Mayhew **BELKNAP (2014)**

Loathed by Karl Marx and admired by Charles Darwin, Enlightenment scholar Thomas Malthus still polarizes, notes historian Robert Mayhew. The flashpoint was Malthus’s 1798 *An Essay on the Principle of Population*, which posits that although humans are prodigal, nature and resources are limited. Mayhew traces that theory through revolutionary and reactionary traditions, arguing that it remains pertinent in an era of economic downturn and shrinking resources, with predictions of 10 billion humans by 2050. **Barbara Kiser**

Correspondence

Carbon cost will not stop oil-sands work

Wendy Palen and colleagues propose a moratorium on new oil-sands projects until regulations are in place to ensure compliance with carbon-emissions commitments (*Nature* **510**, 465–467; 2014). We question whether such a ban is justified on the basis of the criteria they propose.

Emissions from oil-sands production are generally less than 0.1 tonne of carbon dioxide equivalent per barrel, so production costs would increase by at most US\$6.50 per barrel if social costs were accounted for as the authors suggest. The social cost of carbon is estimated at about \$65 per tonne of CO₂ equivalent over the lifespan of an oil-sands project (see go.nature.com/9ztmyu).

However, many projects would remain viable despite this production-cost increase, in part because it would be offset for developers by tax and royalty deductions. The energy company Suncor, for example, estimates that a similar carbon policy would decrease its return on investment on a new mine by just 0.4%.

A moratorium is neither sufficient nor necessary for Canada to meet its greenhouse-gas commitments, or to achieve global carbon stabilization at 450 parts per million (see also N. C. Swart and A. J. Weaver *Nature Clim. Change* **2**, 134–136; 2012).

To meet its target, Canada would probably need to set a carbon price that exceeds the social cost of carbon estimates — say, more than \$100 per tonne (see go.nature.com/dswyma) — and apply it nationally to all sources of emissions. Even then, oil-sands production could continue to grow.

Andrew Leach, Branko Boskovic *University of Alberta, Canada.*

aleach@ualberta.ca
A.L. declares competing financial interests: see go.nature.com/4zleed for details.

Put brain project back on course

As an ambitious initiative of the European Commission, the Human Brain Project (HBP) must unite basic neuroscience research with information and communication technology (see www.humanbrainproject.eu). However, many neuroscientists are concerned that it has failed to do so (see <http://neurofuture.eu>). We believe that it is not too late to put the HBP on course and restore confidence by swift and decisive action.

In our opinion, to let the HBP plough ahead without taking into account widespread views within the European neuroscience community would be akin to giving the lead on climate studies to the critics of global warming. Cooperative and effective large-scale research cannot be decreed: it has to emerge from inclusive discussion and respect for scientific argument.

This will require the European Commission to implement significant changes. For example, the HBP charter needs to be amended to make its governance much more democratic: the direction of a project on this scale must reflect a maximally consensual scientific process. We also suggest that a neuroscience council should be created to formulate a strategy for Europe that is inclusive and scientifically driven, and which would help to drive partnering grants and international collaboration.

We stand committed to working with the European Commission and the HBP, but inclusivity and good governance are essential to meet the huge challenges of understanding the human brain.

Zachary F. Mainen *Champalimaud Neuroscience*

Programme, Lisbon, Portugal.
zmmainen@neuro.fchampalimaud.org
Alexandre Pouget *University Medical Center, University of Geneva, Switzerland.*

Review risks before eradicating toads

Jonathan Kolby and colleagues call for swift eradication of the invasive Asian common toad *Duttaphrynus melanostictus* from Madagascar (*Nature* **509**, 563; 2014). We caution against disproportionate countermeasures that are not founded on proper data and assessment. These could have detrimental effects on local ecosystems that are comparable to the threat posed by the toads themselves.

Draining potential breeding ponds, for example (see *Nature* <http://doi.org/ts3>; 2014), could have an impact on local fauna or even on entire ecosystems. This approach would probably fail anyway because larvae of *D. melanostictus* can survive in streams, puddles and brackish waters. Also, efforts by amateur conservationists and locals to destroy toad spawn and larvae could jeopardize native frog species if people do not identify tadpoles or juveniles correctly (see, for instance, R. Somaweera *et al. Biol. Conserv.* **143**, 1477–1484; 2010).

We consider the parallels drawn by Kolby and colleagues between *D. melanostictus* and the invasive cane toad (*Rhinella marina*) to be inappropriate. Invasion potential and the effects of alien species are hard to predict without sufficient data. To confirm a genuine biological invasion, information first needs to be collected on the toads' range extension and the impact on local flora and fauna.

Before implementing countermeasures, any negative effects should be evaluated. This calls for rapid assessment of the practical difficulties, risks

and prospects of success.
Sven Mecke ** Philipps-Universität Marburg, Germany.*
meckes@staff.uni-marburg.de
**On behalf of 12 correspondents (see go.nature.com/wj2aju for full list).*

University managers misled by metrics

University administrators wishing to arrive at rapid decisions in evaluating staff performance may ignore metrics that seem too sophisticated (*Nature* **510**, 444; 2014, and see J. Adams *Nature* **510**, 470–471; 2014). At King's College London, for example, appallingly blunt metrics are being wielded to determine who should be made redundant.

Some faculty members there are being appraised on grant income or hours of contact teaching, but not both, and without regard to indicators such as publication record, teaching quality or editorial-board membership (see go.nature.com/bbhyjs). These metrics are having a disproportionate effect on staff with both research and teaching commitments — ironically, the university's stated ideal.

Total grant income is in any case a questionable proxy for research quality, and cannot be used to compare the performance of researchers who have different outgoings and funding sources. Examples include basic and medical researchers, or those who work on model organisms that vary markedly in expense. Their grant sizes are unrelated to the quality of their research.

Such misleading measures cannot inform the shrewd decision-making that is essential for tightly funded higher-education management.
Thomas Butts *King's College London, UK.*
thomas.butts@kcl.ac.uk
The author declares competing financial interests: see go.nature.com/u9y5ti for details.

Causes and effects of Antarctic ice

Some 34 million years ago, there was a rapid growth of ice on Antarctica. A modelling study indicates that the ultimate cause of this glaciation was a decrease in the concentration of atmospheric carbon dioxide. [SEE LETTER P.574](#)

DAN LUNT

On page 574 of this issue, Goldner *et al.*¹ tackle a long-standing debate in palaeoclimate science: the causes and effects of the largest climate transition of the past 50 million years, which occurred about 34 million years ago. It was characterized by rapid cooling and growth of Antarctic ice, marking a change from the warm 'greenhouse' climates of the Eocene epoch to the 'icehouse' of the Oligocene epoch.

Using a numerical climate model, the authors investigated the effects of inserting a continental ice sheet onto Antarctica, and found that spatial patterns of ocean cooling predicted by the model agree well with cooling patterns inferred from the geological record of this time period. This cooling, and associated ocean-circulation changes, had previously been attributed to geographical changes in ocean straits and seaways, but Goldner and colleagues conclude that the cooling is better explained as a response to Antarctic ice growth, itself caused by a decrease in the concentration of atmospheric carbon dioxide.

Before the cooling at the Eocene–Oligocene transition (EOT), much of Antarctica was vegetated and was home to flora and fauna that today is found nearer the Equator, including flowering plants, beech forests and marsupials². Two main hypotheses have emerged to explain the cooling and the growth of ice in this period, which occurred over about

300,000 years^{3,4}. The first proposal, called the gateway hypothesis, posits that gradual movements of continental plates over millions of years gradually widened the Drake Passage and Tasman Gateway in the Southern Ocean, allowing increased ocean flow around Antarctica. This led to decreased poleward heat transport, which resulted in cooling of the Antarctic continent and growth of the Antarctic ice sheet. The second proposal, called the CO₂ hypothesis, postulates that a decreased concentration of atmospheric greenhouse gases, in particular CO₂, led directly to cooling of the Antarctic continent and growth of the ice sheet.

One line of evidence previously used in favour of the gateway hypothesis concerns the spatial pattern of EOT ocean-temperature change. This pattern is derived by drilling deep into the modern ocean floor and extracting cores of ancient ocean sediments. The cores are analysed for their chemical and isotopic composition, providing insights into changes in ocean temperature over time. By drilling cores at various locations and depths in the Atlantic, such analysis has revealed increasing cooling down to a depth of 2 kilometres, and increasing cooling from the Equator southwards. Until now, it had been suggested that this signature was best explained by the gateway hypothesis⁵, being similar to that predicted as a response to changes in ocean circulation associated with evolving ocean gateways⁶.

However, Goldner *et al.* suggest that such a

signature could be equally, or better, explained by the CO₂ hypothesis. When the researchers included an enlarged Antarctic ice sheet in their numerical climate model of the EOT, the predicted temperature change in the Atlantic was very similar to that inferred from the ocean sediment cores. This was not the case when they imposed modifications to the ocean gateways in line with the gateway hypothesis (in contrast to previous work⁶, which the authors argue used modelling tools that are less advanced than their own). As such, Goldner *et al.* conclude that their work provides support for the CO₂ hypothesis.

One of the strengths of this paper is that Goldner and colleagues have carefully analysed their model results to understand the climatic mechanisms (Fig. 1) that give rise to the model-predicted patterns of ocean cooling and circulation. However, in my view, several interesting issues still remain.

First, although the authors make inferences about the causes of Antarctic ice-sheet growth, they cannot tackle this explicitly because their modelling does not include a full representation of the interactions between ice and climate. It is possible that a change in gateways caused cooling that led to the growth of the Antarctic ice sheet, and it is these effects that are seen in the geological record. The causes of ice-sheet growth (or retreat) are best understood through the use of integrated climate and ice-sheet models. Ice-sheet models have recently undergone a period of rapid

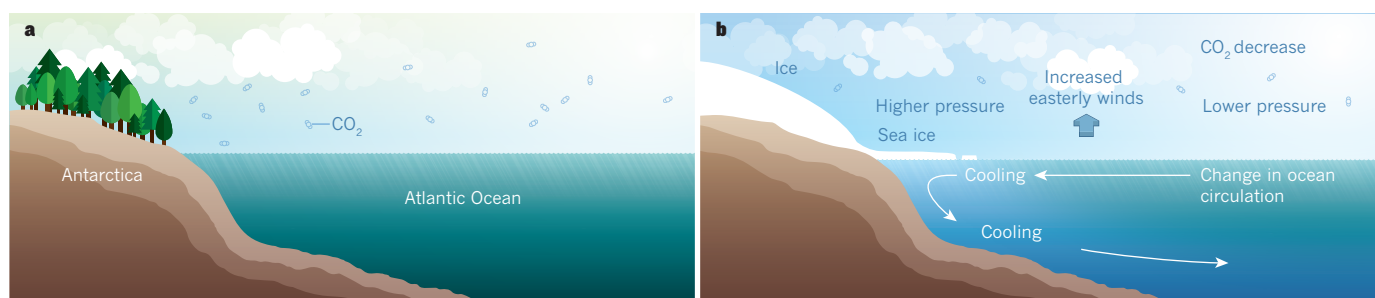


Figure 1 | Earth-system change at the Eocene–Oligocene transition. **a, b,** The diagram shows the Atlantic sector of the high latitudes of the Southern Hemisphere before **(a)** and after **(b)** growth of Antarctic continental ice at the Eocene–Oligocene transition 34 million years ago, as modelled by Goldner and colleagues¹. At this transition, a decreasing concentration of carbon dioxide in the atmosphere leads to atmospheric cooling and growth of Antarctic continental ice and sea ice. The resulting ice sheet induces a north–south atmospheric pressure gradient near the surface, which drives increased easterly surface winds (moving towards the west; into the page as indicated by the blue arrow) around Antarctica. These winds change the ocean circulation, enhancing southwards ocean flow through a process known as Ekman transport. This dense (cold and relatively salty) water mass flows downwards as it reaches the Antarctic coast. Diagram not to scale.

development (see ref. 7, for example), so the time is ripe to apply these to palaeoclimate events such as the EOT.

Second, models are by definition approximations of the real world, and it will be crucial for other groups to verify these findings with their own models. This is particularly important because other recent modelling work⁸ has indicated surface warming in the Atlantic sector in response to an increased Antarctic ice sheet — although that study focused on the more recent Middle Miocene climate transition, which occurred about 14 million years ago. This is in contrast to the earlier cooling found by Goldner and colleagues.

Third, the question also remains as to why greenhouse gases changed at this time, and by how much. Indirect estimates of CO₂

concentration indicate a concurrent drop^{9,10}, although the uncertainties for these estimates are currently large. Possible causes include changes in the balance of sources (for example, decreased volcanism) and/or sinks (such as increased weathering of silicate rocks), and/or changes to reservoirs of carbon (for instance, an increase in the residence time of carbon in the ocean, owing to changes in ocean circulation). Picking apart these possible causes is a crucial challenge.

Goldner and colleagues conclude their paper with a word of warning, noting that a complex web of positive and negative feedbacks means that the climate system can often behave unexpectedly. This can be interpreted as a strong note of caution regarding humanity's own current CO₂ 'experiment' with the climate system. ■

Dan Lunt is in the School of Geographical Sciences, University of Bristol, Bristol BS8 1SS, UK.

e-mail: d.j.lunt@bristol.ac.uk

1. Goldner, A., Herold, N. & Huber, M. *Nature* **511**, 574–577 (2014).
2. Francis, J. E. *et al.* in *Antarctic Climate Evolution* (eds Florindo, F. & Siebert, M.) 309–368 (Elsevier, 2008).
3. Lear, C. H., Bailey, T. R., Pearson, P. N., Coxall, H. K. & Rosenthal, Y. *Geology* **36**, 251–254 (2008).
4. Liu, Z. *et al.* *Science* **323**, 1187–1190 (2009).
5. Cramer, B. S., Toggweiler, J. R., Wright, J. D., Katz, M. E. & Miller, K. G. *Paleoceanography* **24**, PA4216 (2009).
6. Sijp, W. P., England, M. H. & Huber, M. *Paleoceanography* **26**, PA4207 (2011).
7. Cornford, S. L. *et al.* *J. Comput. Phys.* **232**, 529–549 (2013).
8. Knorr, G. & Lohmann, G. *Nature Geosci.* **7**, 376–381 (2014).
9. Pearson, P. N., Foster, G. L. & Wade, B. S. *Nature* **461**, 1110–1113 (2009).
10. Pagani, M. *et al.* *Science* **334**, 1261–1264 (2011).

HIV

The mixed blessing of interferon

A study in monkeys finds that treatment with the protein interferon protects against simian immunodeficiency virus, but that prolonged interferon administration exacerbates the chronic stage of the infection. SEE LETTER P.601

AMALIO TELENTI

Soon after the identification of the AIDS epidemic in the early 1980s, many research groups reported on using interferon, a protein that triggers an antiviral response, to treat HIV-infected patients. However, an early review of the literature¹ concluded that there was no evidence that interferon therapy exerted “any beneficial effect on the underlying immune defects” and that, in any case, many patients with advanced HIV-induced immunosuppression already exhibited elevated levels of interferon before its administration. But on page 601 of this issue, Sandler *et al.*² report intriguing findings on interferons in HIV infection. They show that administering the interferon IFNα2a to rhesus macaques before exposure to simian immunodeficiency virus (SIV) prevented systemic viral infection, whereas treatment with an antagonist of the interferon receptor led to catastrophic infection. But they also find that prolonged exposure to interferon has a detrimental effect.

So what are the exact mediators of interferon that contribute to its beneficial effects in Sandler and colleagues' study? Triggering the interferon response induces the expression of several hundred genes³. The roles of many, if not most, interferon-stimulated proteins are unclear and

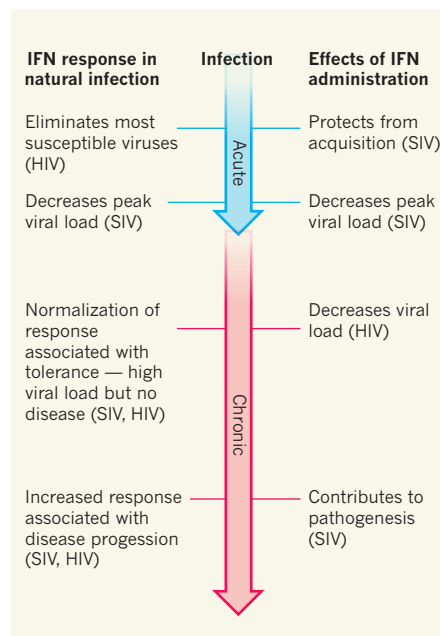


Figure 1 | Interferon in acute and chronic infection. Sandler and colleagues' findings² contribute to other evidence^{4,5,7–9,12} that there are strong parallels between the observed effects of experimentally administered interferon (IFN) on simian immunodeficiency virus (SIV) or HIV infections and the effects of endogenous IFN responses during natural SIV or HIV infections, at both the acute and chronic phases.

little is known about their specificity. Some selectively target HIV and/or SIV, and these are generally referred to as restriction factors⁴. Such factors have protected humans from transmission of retroviruses from other species, but they failed to guard us against SIV from chimpanzees, and HIV is the result of this transmission. Unfortunately, HIV in humans and SIV strains in their natural hosts have developed mechanisms to evade these restriction factors. Thus, Sandler and colleagues' observed success with interferon administration probably results from other aspects of the interferon response, or from the actions of undescribed interferon-stimulated proteins that retain sufficient general or specific antiviral activity. These are not very satisfactory explanations when one considers that the goal is to understand human responses to a therapeutic strategy.

The finding that interferon treatment during SIV challenge increases host resistance to systemic infection reinforces previous data from studies in non-human primates, and places in perspective some existing data on interferon administration in humans (Fig. 1). In HIV-infected individuals⁵ and in individuals co-infected with hepatitis C virus⁶, use of interferon results in a modest decrease in HIV viral load in the blood. It was therefore a logical step for Sandler *et al.* to assess the effects of prolonged interferon administration during SIV infection. They observed that, despite the early beneficial effects, continued administration resulted in increased susceptibility to infection and greater depletion of CD4⁺ T cells (immune cells that are killed by SIV and HIV), and decreased expression of interferon-stimulated genes, compared to placebo.

They describe this paradox as an interferon-desensitized state, and demonstrate that this is not the result of the animals developing neutralizing antibodies against the interferon. Interestingly, earlier studies had described a refractory state of cells to repeat interferon induction⁷. Furthermore, there is evidence^{8,9}

that a persistent abnormal interferon response — elevated expression of interferon-stimulated genes — is linked to increased viral load and disease severity during the chronic phase of HIV infection. Thus, both continuous administration of interferon and persistent elevation of interferon-stimulated genes are associated with unfavourable outcomes in chronic HIV and SIV infection.

HIV is not thought to become resistant to interferon in chronic infection — in fact, recent data suggest the contrary¹⁰. Instead, there is increasing concern that when the host cannot clear an infection, continued interferon signalling leads to the induction of immuno-suppressive pathways with the aim of limiting damage associated with chronic infection (reviewed in ref. 11). This concept resonates with the idea of ‘tolerance’, in which the host attempts to reduce the negative impact of an infection without directly affecting pathogen burden¹². Tolerance is frequently observed in non-pathogenic SIV infections in natural hosts — animals may have a high viral load but there is minimal pathogenesis, and the interferon response normalizes after the initial stage of infection⁹. Unfortunately, in HIV infections the host rarely finds the sweet spot between viral load, disease and interferon: tolerance is rarely observed in humans¹³ and progressive immunosuppression almost always ensues.

Overall, there are strong parallels between the effects of deliberate interferon administration in experimental SIV infections and the outcome of endogenous interferon responses in natural HIV and SIV infection (Fig. 1). Sandler and colleagues’ research highlights the importance of the timing and duration of interferon administration, but also underscores the difficulty of understanding the exact potential of intervening in the interferon signalling pathways during infection.

The interferon response, and more broadly, the innate immune response, remains a field of unknowns¹⁴. Challenges include identifying the exact set of effector molecules, their roles at portals of entry of pathogens and their individual actions in acute and chronic disease. Thus, deconvoluting the interferon response is needed if the aim is to better use this pathway for therapeutic purposes. The complex antiviral response to this protein can be summarized by the Cantonese expression ‘equipped with knives all over, yet none is sharp’ — a fitting metaphor for the failure of interferon to cut cleanly through HIV infections. ■

Amalio Telenti is at the Institute of Microbiology, University Hospital of Lausanne, University of Lausanne, 1011 Lausanne, Switzerland.
e-mail: amalio.telenti@chuv.ch

1. Longo, D. L. *et al.* *Ann. NY Acad. Sci.* **437**, 421–430 (1984).
2. Sandler, N. G. *et al.* *Nature* **511**, 601–605 (2014).
3. Schoggings, J. W. *Curr. Opin. Virol.* **6C**, 40–46 (2014).

4. Harris, R. S., Hultquist, J. F. & Evans, D. T. *J. Biol. Chem.* **287**, 40875–40883 (2012).
5. Asmuth, D. M. *et al.* *J. Infect. Dis.* **201**, 1686–1696 (2010).
6. Torriani, F. J. *et al.* *N. Engl. J. Med.* **351**, 438–450 (2004).
7. Borden, E. C. & Murphy, F. A. *J. Immunol.* **106**, 134–142 (1971).
8. Rotger, M. *et al.* *PLoS Pathog.* **6**, e1000781 (2010).
9. Bosinger, S. E. *et al.* *J. Clin. Invest.* **119**, 3556–3572 (2009).

10. Fenton-May, A. E. *et al.* *Retrovirology* **10**, 146 (2013).
11. Ivashkiv, L. B. & Donlin, L. T. *Nature Rev. Immunol.* **14**, 36–49 (2014).
12. Medzhitov, R., Schneider, D. S. & Soares, M. P. *Science* **335**, 936–941 (2012).
13. Rotger, M. *et al.* *J. Clin. Invest.* **121**, 2391–2400 (2011).
14. Rausell, A., McLaren, P. J. & Telenti, A. *F1000Prime Rep.* **5**, 29 (2013).

This article was published online on 9 July 2014.

QUANTUM PHYSICS

The path most travelled

Continuous tracking of the random trajectories of a superconducting quantum system as it evolves between two selected initial and final states has allowed researchers to determine the most probable path of the system. [SEE LETTER P.570](#)

ADRIAN LUPASCU

Trajectories of various kinds, such as those of aeroplanes and migrating birds, are a familiar part of our everyday experience. A trajectory is given, classically, by the position coordinates of an object as a function of time. In the realm of quantum mechanics, the state of an object is given not in terms of position, but rather by a more abstract mathematical construct — a vector in the Hilbert space. Nevertheless, quantum states can be

parameterized by a set of coordinates whose evolution in time defines quantum trajectories. For a quantum system in isolation, quantum trajectories are deterministic, bearing a deceptive similarity to classical trajectories. However, the act of observation makes quantum trajectories random, revealing their fragile character. On page 570 of this issue, Weber *et al.*¹ analyse these random trajectories and find that the most likely ones still provide insightful information about a quantum system.

To study quantum trajectories, Weber and

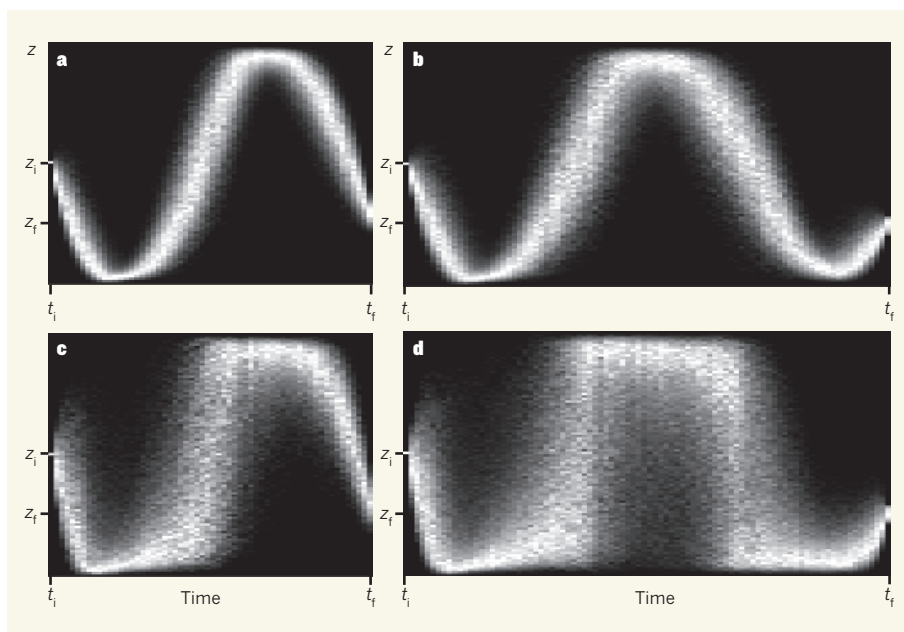


Figure 1 | Transmon trajectories. Weber *et al.*¹ have measured the evolution of the quantum state of a transmon superconducting device. The quantum state is represented by the coordinate z , and its change over time defines a quantum trajectory. In the four sets of trajectories shown here (a–d), trajectories were selected that had initial and final coordinates, at times t_i and t_f , around z_i and z_f respectively. The four panels correspond to four experiments that differ in the strength of the measurement (c and d correspond to stronger measurements) and in the total evolution time (b and d correspond to a longer evolution time). The trajectories are ‘blurred’, reflecting the randomness of the measurements and the corresponding random change of state. However, the average trajectory in each case carries information on the dynamics of the system.

colleagues used a transmon, a micrometre-sized superconducting device that at low temperatures behaves as a quantum two-state system, with states dubbed 0 and 1. The transmon and other superconducting quantum devices are under intense investigation because of their potential applications in quantum computing² and as test systems for investigating fundamental aspects of quantum mechanics³.

The authors' experiment builds on major advances in this field over the past few years. The first of these is the reduction in the 'decoherence' of transmons. Any quantum system that interacts with its environment will 'forget' its state — that is, the system will decohere. In general, this effect hampers the observation of quantum behaviour⁴ and would mask the quantum effects observed by Weber and colleagues. The second advance is the development of 'almost-perfect' measurement methods for transmons. In quantum mechanics, even perfect measurements are fundamentally limited: any single measurement provides only partial information on the state of an observed system. Moreover, after a measurement has been taken, the state of the system is changed irreversibly. The key feature of perfect quantum measurements is that the change of state following a measurement is the smallest possible allowed by the fundamental laws of quantum mechanics⁵.

Weber *et al.* measured the state of the transmon by coupling it to a device known as a superconducting resonator. The natural frequency of oscillation of microwaves in this resonator depends on the transmon's state. The authors sent microwaves to the resonator and continuously monitored how they were scattered. The scattered microwaves were processed further to extract a continuous signal containing information related to the state of the transmon.

This detection scheme involves a continuous 'weak' measurement of the transmon. Weak measurements have been investigated both theoretically⁶ and experimentally⁷. To understand this type of measurement, assume that, before a measurement is started, the quantum state of the system under investigation is well known. Next, the measurement apparatus is turned on. The measurement signal taken over a short time interval is a nearly random quantity, by itself insufficient to infer the system's state. However, the prior knowledge of the state combined with the tiny bit of information obtained from the measured signal is enough to fully infer the system's new quantum state. This process can be extended over the full duration of the measurement procedure; by using the continuous signal from the measurement apparatus, knowledge of the quantum state can be continuously updated. The change of the state over time is a quantum trajectory.

The random nature of quantum trajectories reflects the fact that the change of the quantum state at each time in the measurement process depends on the measurement result, which is

itself random. In their experiment, Weber *et al.* analyse trajectories in the following way. A set of trajectories is selected that is conditional on the initial and final states of the transmon. Although such a set of trajectories is random, the most likely one is found to provide valuable information about the transmon (Fig. 1). The most probable path is one that reflects, on the one hand, the tendency of the transmon to settle in state 0 or 1 and, on the other hand, its tendency to oscillate between these two states.

The most likely trajectory can be theoretically calculated⁸ by requiring that a global measure of the trajectory, the action, is an extremum — that is, insensitive with respect to small changes in the trajectory. This approach establishes an intriguing connection with other theories in which path optimization is key, such as Fermat's least-time principle for light propagation, the Hamilton principle for dynamics in classical mechanics, and also the formulation of quantum mechanics in terms of mathematical entities known as path integrals.

Weber *et al.* have successfully measured the statistics of quantum trajectories for their transmon device and shown that the most probable trajectory is in agreement with calculations based on extremal action⁸. An interesting parallel may be drawn with classical trajectories. Fluorescent markers can be used to characterize flow patterns or biological processes. Analogously, quantum trajectories carry information about the time dynamics of quantum systems. The use of weak measurements to determine quantum trajectories can therefore provide information about the parameters of the system that generate the dynamics. Another potential application of weak measurements is the preparation of quantum states.

Further development of the experiments described here will have to address the fidelity of the measurement procedure, which has an efficiency of 40% in its current form. The use of this method for quantum parameter and state estimation will require rigorous investigation, particularly with regard to how the method compares with similar protocols based on strong quantum measurements. ■

Adrian Lupascu is in the Department of Physics and Astronomy and the Institute for Quantum Computing, University of Waterloo, Waterloo, Ontario N2L 3G1, Canada.
e-mail: adrian.lupascu@uwaterloo.ca

1. Weber, S. J. *et al.* *Nature* **511**, 570–573 (2014).
2. Devoret, M. H. & Schoelkopf, R. J. *Science* **339**, 1169–1174 (2013).
3. You, J. Q. & Nori, F. *Nature* **474**, 589–597 (2011).
4. Haroche, S. & Raimond, J. M. *Exploring the Quantum: Atoms, Cavities, and Photons* (Oxford Univ. Press, 2006).
5. Braginsky, V. B. & Khalili, F. Ya. *Quantum Measurement* (Cambridge Univ. Press, 1995).
6. Korotkov, A. N. *Phys. Rev. B* **60**, 5737 (1999).
7. Murch, K. W., Weber, S. J., Macklin, C. & Siddiqi, I. *Nature* **502**, 211–214 (2013).
8. Chantasri, A., Dressel, J. & Jordan, A. N. *Phys. Rev. A* **88**, 042110 (2013).



50 Years Ago

At the meeting of the Society for Visiting Scientists on June 3, great interest was expressed in the potentialities of international research centres. It was admitted from the outset that there is no intrinsic merit in international research as such, since the important thing in any scientific work is the result, not how or where the work is carried out. The justification for any proposed international effort must therefore be carefully examined ... The experience of running CERN presented some interesting lessons ... CERN had been formed at a time when there was a sense of togetherness among most of the nations of Western Europe, a feeling which sought for some practical expression. It was important that any concrete form which could be given to it should not be controversial, should not be military, should not be a political disaster if it failed and that success if achieved should be clearly recognizable. If these factors were present, the way was open for the scientists, who, after all, led the world in international co-operation, to exert pressure on their political colleagues.
From Nature 1 August 1964

100 Years Ago

Many instances are on record of so-called "wolf-children," said to have been found in the jungles of India. A strange story is now reported from Naini Tal, the summer capital of the United Provinces of Agra and Oudh, of a female child about nine years old found in this neighbourhood, and unable to eat anything except grass and chapatis or native griddle cakes. She has a great mat of head hair and a thick growth on the sides of her face and spine. She bears marks of vaccination and is clearly a child who had, years ago, been abandoned or strayed into the jungle.
From Nature 30 July 1914

EPIGENETICS

Cellular memory erased in human embryos

Two analyses of human eggs, sperm and early-stage embryos reveal a pronounced loss of DNA methylation — a molecular modification that affects gene transcription — after fertilization. [SEE LETTERS P.606 & P.611](#)

WOLF REIK & GAVIN KELSEY

Epigenetic modifications are changes to the genome that can affect gene expression without altering DNA sequence. Like DNA itself, certain epigenetic modifications can be copied faithfully when cells divide, allowing daughter cells to retain this information from their parents. This ensures that gene expression is maintained in a stable manner down cell lineages. One such modification is methylation, whereby methyl groups are added to DNA. Two papers in this issue^{1,2} show that there is a massive loss of DNA methylation from most of the genome immediately after fertilization in human embryos. Thus, methylation memory is erased on a global scale — an epigenetic reprogramming step that seems to be fundamental in mammals.

DNA methylation usually represses transcription, and primarily occurs on cytosine bases in the dinucleotide sequence cytosine-guanine (CpG, where p denotes the phosphate backbone of DNA, indicating that the nucleotides are on the same DNA strand). Because Watson-Crick base-pairing dictates that C pairs with G on complementary DNA strands, CpG sequences align and both strands

are methylated in the same place. Therefore, methylation patterns can be passed on when cells divide, through the CpG 'memory module'. This inheritance of epigenetic information is vital in specialized cell lineages, which must maintain their identity as they divide — for example, dividing blood cells maintain their epigenetic identity to give rise to daughters that are also blood cells.

Guo *et al.*¹ (page 606) and Smith *et al.*² (page 611) analysed genome-wide DNA methylation in early-stage human embryos by high-throughput sequencing. They studied eggs, sperm, fertilized eggs (zygotes) and embryos at various stages of development, including the blastocyst stage, which occurs just before the embryo becomes implanted in the uterus, and a post-implantation stage. Both groups found that the DNA of sperm was highly methylated and that of eggs moderately so (much like mouse sperm and eggs^{3–5}). However, zygotes and two-cell embryos had lost a large proportion of this methylation. In particular, Guo *et al.* observed marked demethylation of the paternal, sperm-derived genome, compared with more-modest demethylation of the maternal genome.

At the blastocyst stage, methylation levels

remained low. This was true in all blastocyst cell types, including the cells of a structure called the inner cell mass, which are pluripotent — they can give rise to every cell of the body. Previous research indicates⁶ that epigenetic memory must be erased for embryonic cells to achieve pluripotency, providing a possible explanation for global demethylation. By contrast, both groups observed that, after implantation, when cells had begun to adopt tissue-specific identities, DNA methylation rapidly rose to a level characteristic of differentiated cells. After a near-total wipe-out, the epigenetic memory system was back in place (Fig. 1).

These results, combined with those from mice^{7,8} and other mammals^{9,10}, suggest that global methylation reprogramming after fertilization is evolutionarily conserved. Perhaps this is because early-stage mammalian embryos undergo rapid transcriptional activation, together with early diversification of cell types — factors that necessitate a transient pluripotent state. Nonetheless, it is remarkable that the demethylation kinetics of mouse and human embryos are so similar, given that other aspects of their early development are less conserved. For example, the major transcriptional activation of the embryonic genome occurs at the two-cell stage in mouse embryos, whereas in humans it takes place at the transition between four and eight cells.

It is exciting that Guo and colleagues detected an alternative form of epigenetic modification called hydroxymethylation preferentially in the paternal genome, because hydroxymethylation is implicated in demethylation in mice^{11–13}. This reinforces the idea that major mechanisms of epigenetic reprogramming are conserved in mammals. The studies did not address the mechanisms of demethylation further. Such analyses are

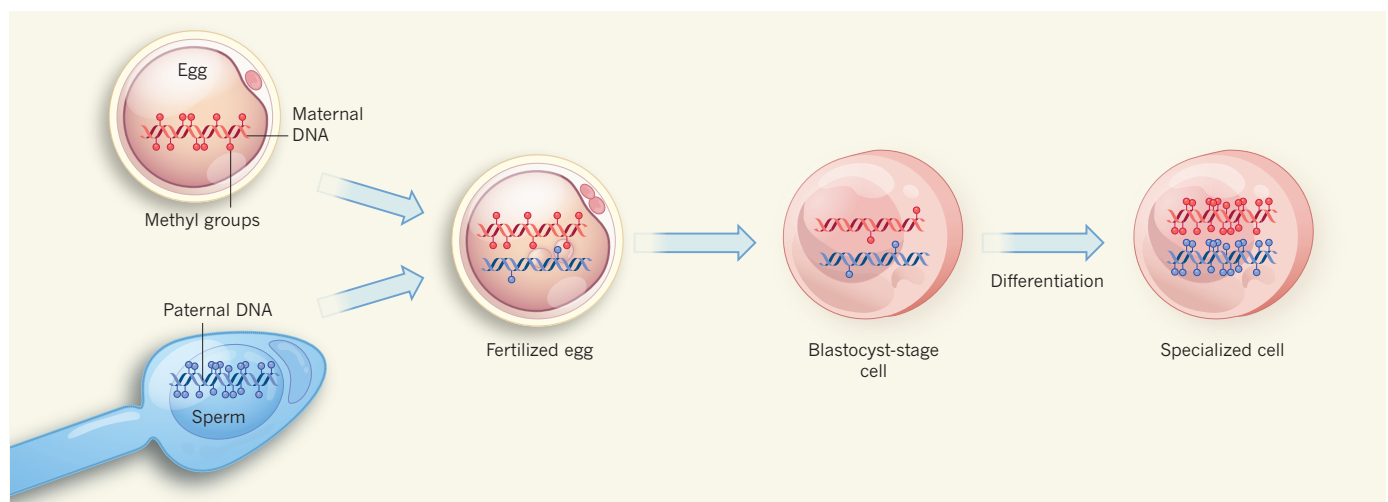


Figure 1 | Tracking the state of DNA methylation. Guo *et al.*¹ and Smith *et al.*² investigated DNA methylation during early development of the human embryo. The DNA of human sperm is highly methylated, and that of eggs less so (sperm and egg not drawn to scale). However, once the egg has been fertilized, methylation is largely lost — more so from the paternal than from

the maternal genome. As the embryo begins to develop, methylation marks continue to be lost from the maternal genome of cells up to the blastocyst stage. After this stage, DNA in differentiating cells becomes remethylated, allowing specialized cell types to pass instructions about control of gene transcription to their daughters.

challenging in human embryos, but Smith and co-workers have taken a first step, growing pluripotent embryonic stem cells derived from blastocyst-stage embryos *in vitro*, and finding that the cells become rapidly remethylated. This might be a viable system for manipulating and so studying genome-wide methylation and demethylation in human embryos, as is possible in mice.

Genome-wide analyses permit a detailed survey of distinct regions of DNA sequence whose function is known to be modified by methylation, allowing investigation of how they behave in the face of global demethylation. Such regions include 'imprinted' genes, CpG-rich genetic regions called CpG islands, and transposons (DNA sequences that can move about the genome).

Imprinted genes are those that are expressed preferentially from one parental chromosome (maternal or paternal), unlike most genes, which can be expressed from both chromosomes. Unusually, epigenetic memory in imprinted sequences is retained throughout development. The two groups confirmed this in human embryos, which they found carried methylation memories from the embryos' parents in conserved imprinted regions.

The authors found that, in contrast to sperm, human eggs had hundreds of methylated CpG islands that differed from those in

mouse eggs^{3–5} and, as a general rule, these maternal epigenetic marks were not well maintained after fertilization in the embryos of mice¹⁴ or humans. Perhaps this reflects a difference in the development of the egg in the two species that is no longer relevant after fertilization. Alternatively, some of these maternal epigenetic signals may be required only in the early embryo, and thus could contribute to species differences in imprinting, particularly in the placenta¹⁵.

Transposons need to be treated with caution during reprogramming, because demethylation might cause their transcriptional activation. If they are evolutionarily 'young' and relatively unmutated, this might lead to their being able to move around in the genome, which could result in unwanted mutations. Guo and colleagues investigated one class of transposon, LINE elements, and found that evolutionarily young elements were more resistant to demethylation than their older counterparts.

The new studies provide an atlas of methylation reprogramming in early human embryos and hence a foundation for studying epigenetic regulation of human development. This is vital if we are to understand the epigenetic mechanisms that control pluripotency and differentiation. Such understanding will also help in assessing the long-term consequences

of fertility interventions, including *in vitro* fertilization, for human health. ■

Wolf Reik and Gavin Kelsey are in the Epigenetics Programme, Babraham Institute, Babraham Research Campus, Cambridge CB22 3AT, UK, and at the Centre for Trophoblast Research, University of Cambridge. W.R. is also at the Wellcome Trust Sanger Institute, Cambridge.

1. Guo, H. *et al.* *Nature* **511**, 606–610 (2014).
2. Smith, Z. D. *et al.* *Nature* **511**, 611–615 (2014).
3. Smallwood, S. A. *et al.* *Nature Genet.* **43**, 811–814 (2011).
4. Kobayashi, H. *et al.* *PLoS Genet.* **8**, e1002440 (2012).
5. Smith, Z. D. *et al.* *Nature* **484**, 339–344 (2012).
6. Lee, H. J., Hore, T. A. & Reik, W. *Cell Stem Cell* **14**, 710–719 (2014).
7. Mayer, W., Niveleau, A., Walter, J., Fundele, R. & Haaf, T. *Nature* **403**, 501–502 (2000).
8. Santos, F., Hendrich, B., Reik, W. & Dean, W. *Dev. Biol.* **241**, 172–182 (2002).
9. Dean, W. *et al.* *Proc. Natl Acad. Sci. USA* **98**, 13734–13738 (2001).
10. Fulka, H., Mrazek, M., Tepla, O. & Fulka, J. Jr *Reproduction* **128**, 703–708 (2004).
11. Iqbal, K., Jin, S.-G., Pfeifer, G. P. & Szabó, P. E. *Proc. Natl Acad. Sci. USA* **108**, 3642–3647 (2011).
12. Wossidlo, M. *et al.* *Nature Commun.* **2**, 241 (2011).
13. Gu, T.-P. *et al.* *Nature* **477**, 606–610 (2011).
14. Proudhon, C. *et al.* *Mol. Cell* **47**, 909–920 (2012).
15. Court, F. *et al.* *Genome Res.* **24**, 554–569 (2014).

This article was published online on 23 July 2014.

MALARIA

Protein-export pathway illuminated

Two studies provide evidence that the protein complex PTEX is needed for export of malaria-parasite proteins into the cytoplasm of infected cells, and that such export is essential for parasite survival. [SEE LETTERS P.587 & P.592](#)

SANJAY A. DESAI & LOUIS H. MILLER

Malaria parasites export hundreds of proteins into the red blood cells that they infect. These proteins increase nutrient uptake from blood plasma, facilitate adhesion of the infected cell to endothelial cells in blood vessels and markedly remodel the red blood cell for the parasite's benefit. A parasite protein complex called *Plasmodium* translocon of exported proteins (PTEX) has been proposed to traffic these proteins across the membrane of the cellular vacuole that separates the parasite from the cytoplasm of the infected cell¹. In two papers in this issue, Beck *et al.*² (page 592) and Elsworth *et al.*³ (page 587) definitively show protein transport through PTEX.

Protein interaction studies¹ have suggested

that the PTEX translocon consists of five proteins and that one of these, EXP2, forms a pore through which proteins are threaded after they have been unfolded, in an energy-dependent process, by the chaperone protein HSP101. The new studies demonstrate the function of this protein complex in the infected cell by ablating the activity of HSP101 and PTEX150, a PTEX component with unknown function. Suppression of either component, which was achieved by transcriptional repression or protein destabilization, was found to prevent export of two broad categories of parasite protein.

In *Plasmodium falciparum*, one of the main human-infecting malaria parasites, most exported proteins contain a sequence of five amino-acid residues, called the PEXEL motif, near their amino terminus^{4,5}. Cleavage within

this motif by the enzyme plasmepsin V commits the mature protein for export to the host cell, but the precise nature of the commitment step remains debated⁶. The second category of exported proteins lacks this motif. Such PEXEL-negative exported proteins, of which there is an expanding list, lack unifying features and thus have defied attempts to predict how they are recognized and exported⁷. The authors' observations of inhibited export of proteins from both categories — which cover the full spectrum of export timings, protein biophysical properties and destinations in the host cell — implicates PTEX as a crucial bottleneck in parasite-induced remodelling of the red blood cell.

These findings raise new questions about the translocation process. How does PTEX recognize and transport such diverse proteins, but allow other proteins to remain in the parasitophorous vacuole (the parasite-containing compartment in the host cell that is formed during invasion and further modified during parasite growth)? Are retained proteases, chaperones and other enzymes not recognized by PTEX or do they carry specific signals that prevent export? To be threaded through PTEX, exported proteins must first be unfolded, as has been shown with a reporter protein that was prohibited from unfolding by a tightly bound substrate⁸. Whether HSP101 or other chaperones actually catalyse the unfolding of each exported protein is unknown. Also still

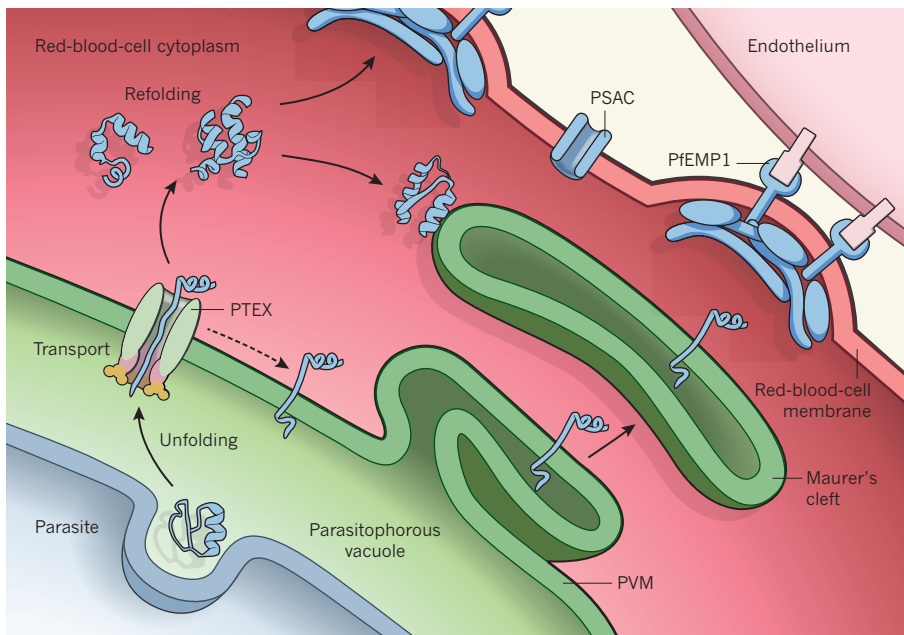


Figure 1 | Transport of parasite proteins into the host red blood cell. The protein complex PTEX mediates export of malaria-parasite proteins across the parasitophorous vacuolar membrane (PVM), which separates the cytoplasm of an infected cell from the vacuole in which the parasite resides. Parasite proteins are first secreted into the vacuolar space, in which they are unfolded before transport through PTEX; they are then refolded in the host-cell cytoplasm. Some membrane proteins might undergo lateral transfer (dashed arrow) from PTEX into the PVM, a process that would allow movement within membranes to reach vesicles known as Maurer's clefts. Exported proteins then localize to specific sites in the cell or on the red-cell membrane, at which they serve functions that are crucial to intracellular parasite growth. Examples include nutrient uptake through PSAC and PfEMP1-mediated binding of infected cells to endothelial cells that line blood vessels.

unclear are if EXP2 indeed defines the pore and what roles the other PTEX components might serve. Could membrane proteins passing through PTEX undergo lateral transfer into the parasitophorous vacuolar membrane to allow migration along membranous extensions (Fig. 1), as established for translocons in other organisms⁹? Finally, proteins that pass into the host-cell cytoplasm will require refolding, presumably by parasite chaperones that are also exported and must somehow be refolded themselves¹⁰.

Another fundamental finding of these studies is that suppression of protein export interferes with intracellular parasite growth, indicating that exported proteins have essential roles in parasite survival. The authors observed adverse effects on parasite development *in vitro* and *in vivo*, with immature ring-stage parasites unable to mature to the trophozoite stage. By contrast, inhibiting PTEX after maturation to the trophozoite stage was well tolerated, with no effect on parasite egress from the cell or invasion of new red blood cells, suggesting that these latter processes do not depend on proteins exported late in the cycle. But development of early-stage gametocytes, the sexual stage of the parasite life cycle required for malaria transmission by mosquitoes, was also severely compromised.

Which activities of the numerous exported proteins account for the parasite growth inhibition seen in these studies? Although

binding of infected cells to endothelial receptors is required for parasite survival *in vivo*, it is dispensable for *in vitro* culture. A leading candidate is the uptake of nutrients by the plasmodial surface anion channel (PSAC), an essential activity associated with the parasite protein cytoadherence-linked antigen 3 (CLAG3)^{11,12}. Beck and colleagues found that CLAG3 still enters the host-cell cytoplasm when PTEX is suppressed, implying that it is exported by a distinct mechanism, perhaps during invasion. At the same time, solute transport by PSAC was curtailed, suggesting that other exported proteins are required for nutrient-channel formation.

In human malaria, binding of infected cells to the endothelium averts their destruction by the spleen and is primarily mediated by members of the *P. falciparum* erythrocyte membrane protein 1 (PfEMP1) protein family. Each member has multiple binding domains at its extracellular face and a single transmembrane domain to anchor the protein over parasite-induced knobs on the infected cell¹³. How does PfEMP1 move from the parasite to the red-blood-cell membrane? Although the protein is not cleaved by plasmepsin V, its atypical PEXEL motif and transmembrane domain both seem to contribute to its export⁴. Subsequent refolding of the endothelium-binding domains in the host cytoplasm presumably requires disulphide-isomerase enzymes to bring numerous cysteine

amino-acid residues together correctly and may involve a battery of chaperones^{14,15}. Specialized sorting organelles known as Maurer's clefts and proteins at the surface knobs also seem to be required for the ultimate insertion of PfEMP1 in the host membrane¹⁶. In light of the complex folded structure of PfEMP1 and the possible involvement of many chaperones, the compromised export of this protein observed by both Beck *et al.* and Elsworth *et al.* could reflect an indirect effect of PTEX inhibition. Further study will be required to determine the precise mechanisms for trafficking and presentation of this key virulence factor.

The two new articles reveal a remarkably broad range of substrates for the translocon and provide compelling evidence that protein export is essential for the parasite and therefore represents a potential therapeutic target. We foresee that combinations of drugs that target both PTEX and exported parasite activities, such as PSAC-mediated nutrient uptake, may be highly synergistic antimalarial therapies. ■

Sanjay A. Desai and Louis H. Miller are in the Laboratory of Malaria and Vector Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Rockville, Maryland 20852, USA. e-mails: sdesai@niaid.nih.gov; lmiller@niaid.nih.gov

1. de Koning-Ward, T. F. *et al. Nature* **459**, 945–949 (2009).
2. Beck, J. R., Muralidharan, V., Oksman, A. & Goldberg, D. E. *Nature* **511**, 592–595 (2014).
3. Elsworth, B. *et al. Nature* **511**, 587–591 (2014).
4. Marti, M., Good, R. T., Rug, M., Knuepfer, E. & Cowman, A. F. *Science* **306**, 1930–1933 (2004).
5. van Ooij, C. *et al. PLoS Pathogens* **4**, e1000084 (2008).
6. Goldberg, D. E. & Cowman, A. F. *Nature Rev. Microbiol.* **8**, 617–621 (2010).
7. Heiber, A. *et al. PLoS Pathogens* **9**, e1003546 (2013).
8. Gehde, N. *et al. Mol. Microbiol.* **71**, 613–628 (2009).
9. Van den Berg, B. *et al. Nature* **427**, 36–44 (2004).
10. Pesce, E.-R. & Blatch, G. L. *Parasitology* <http://dx.doi.org/10.1017/S003118201300228X> (2014).
11. Nguiragool, W. *et al. Cell* **145**, 665–677 (2011).
12. Pillai, A. D. *et al. Mol. Pharmacol.* **82**, 1104–1114 (2012).
13. Juillerat, A. *et al. Proc. Natl Acad. Sci. USA* **108**, 5243–5248 (2011).
14. Maier, A. G. *et al. Cell* **134**, 48–61 (2008).
15. K  lzer, S. *et al. Cell. Microbiol.* **14**, 1784–1795 (2012).
16. Papakrivou, J., Newbold, C. I. & Lingelbach, K. *Mol. Microbiol.* **55**, 1272–1284 (2005).

This article was published online on 16 July 2014.

CORRECTION

The final corrections (to remove mentions of specific chromosomes) to the Retractions (*Nature* **511**, 112; 2014) were accidentally omitted from the print versions. The online versions were correct.

Comprehensive molecular profiling of lung adenocarcinoma

The Cancer Genome Atlas Research Network*

Adenocarcinoma of the lung is the leading cause of cancer death worldwide. Here we report molecular profiling of 230 resected lung adenocarcinomas using messenger RNA, microRNA and DNA sequencing integrated with copy number, methylation and proteomic analyses. High rates of somatic mutation were seen (mean 8.9 mutations per megabase). Eighteen genes were statistically significantly mutated, including *RIT1* activating mutations and newly described loss-of-function *MGA* mutations which are mutually exclusive with focal *MYC* amplification. *EGFR* mutations were more frequent in female patients, whereas mutations in *RBM10* were more common in males. Aberrations in *NF1*, *MET*, *ERBB2* and *RIT1* occurred in 13% of cases and were enriched in samples otherwise lacking an activated oncogene, suggesting a driver role for these events in certain tumours. DNA and mRNA sequence from the same tumour highlighted splicing alterations driven by somatic genomic changes, including exon 14 skipping in *MET* mRNA in 4% of cases. MAPK and PI(3)K pathway activity, when measured at the protein level, was explained by known mutations in only a fraction of cases, suggesting additional, unexplained mechanisms of pathway activation. These data establish a foundation for classification and further investigations of lung adenocarcinoma molecular pathogenesis.

Lung cancer is the most common cause of global cancer-related mortality, leading to over a million deaths each year and adenocarcinoma is its most common histological type. Smoking is the major cause of lung adenocarcinoma but, as smoking rates decrease, proportionally more cases occur in never-smokers (defined as less than 100 cigarettes in a lifetime). Recently, molecularly targeted therapies have dramatically improved treatment for patients whose tumours harbour somatically activated oncogenes such as mutant *EGFR*¹ or translocated *ALK*, *RET*, or *ROS1* (refs 2–4). Mutant *BRAF* and *ERBB2* (ref. 5) are also investigational targets. However, most lung adenocarcinomas either lack an identifiable driver oncogene, or harbour mutations in *KRAS* and are therefore still treated with conventional chemotherapy. Tumour suppressor gene abnormalities, such as those in *TP53* (ref. 6), *STK11* (ref. 7), *CDKN2A*⁸, *KEAP1* (ref. 9), and *SMARCA4* (ref. 10) are also common but are not currently clinically actionable. Finally, lung adenocarcinoma shows high rates of somatic mutation and genomic rearrangement, challenging identification of all but the most frequent driver gene alterations because of a large burden

of passenger events per tumour genome^{11–13}. Our efforts focused on comprehensive, multiplatform analysis of lung adenocarcinoma, with attention towards pathobiology and clinically actionable events.

Clinical samples and histopathologic data

We analysed tumour and matched normal material from 230 previously untreated lung adenocarcinoma patients who provided informed consent (Supplementary Table 1). All major histologic types of lung adenocarcinoma were represented: 5% lepidic, 33% acinar, 9% papillary, 14% micropapillary, 25% solid, 4% invasive mucinous, 0.4% colloid and 8% unclassifiable adenocarcinoma (Supplementary Fig. 1)¹⁴. Median follow-up was 19 months, and 163 patients were alive at the time of last follow-up. Eighty-one percent of patients reported past or present smoking. Supplementary Table 2 summarizes demographics. DNA, RNA and protein were extracted from specimens and quality-control assessments were performed as described previously¹⁵. Supplementary Table 3 summarizes molecular estimates of tumour cellularity¹⁶.

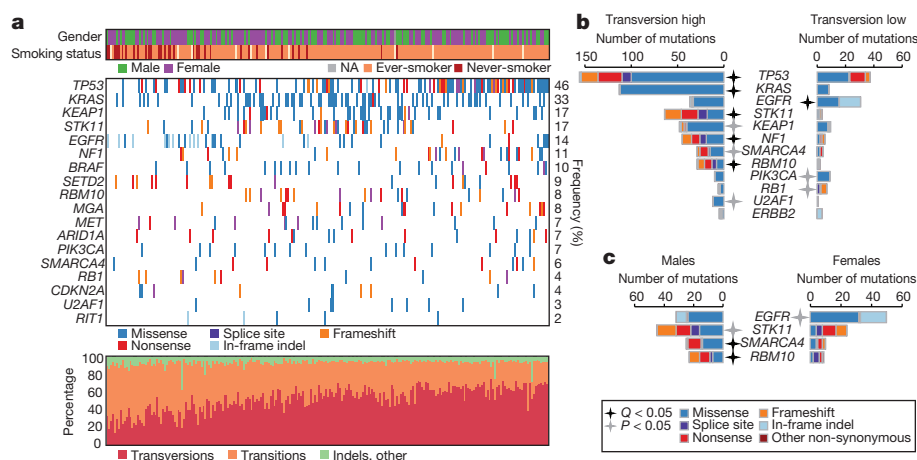


Figure 1 | Somatic mutations in lung adenocarcinoma. **a**, Co-mutation plot from whole exome sequencing of 230 lung adenocarcinomas. Data from TCGA samples were combined with previously published data¹² for statistical analysis. Co-mutation plot for all samples used in the statistical analysis ($n = 412$) can be found in Supplementary Fig. 2. Significant genes with a corrected P value less than 0.025 were identified using the MutSig2CV algorithm and are ranked in order of decreasing prevalence. **b**, **c**, The differential patterns of mutation between samples classified as transversion high and transversion low samples (**b**) or male and female patients (**c**) are shown for all samples used in the statistical analysis ($n = 412$). Stars indicate statistical significance using the Fisher's exact test (black stars: $q < 0.05$, grey stars: $P < 0.05$) and are adjacent to the sample set with the higher percentage of mutated samples.

*A list of authors and affiliations appears at the end of the paper.

Somatically acquired DNA alterations

We performed whole-exome sequencing (WES) on tumour and germline DNA, with a mean coverage of 97.6× and 95.8×, respectively, as performed previously¹⁷. The mean somatic mutation rate across the TCGA cohort was 8.87 mutations per megabase (Mb) of DNA (range: 0.5–48, median: 5.78). The non-synonymous mutation rate was 6.86 per Mb. MutSig2CV¹⁸ identified significantly mutated genes among our 230 cases along with 182 similarly-sequenced, previously reported lung adenocarcinomas¹². Analysis of these 412 tumour/normal pairs highlighted 18 statistically significant mutated genes (Fig. 1a shows co-mutation plot of TCGA samples ($n = 230$), Supplementary Fig. 2 shows co-mutation plot of all samples used in the statistical analysis ($n = 412$) and Supplementary Table 4 contains complete MutSig2CV results, which also appear on the TCGA Data Portal along with many associated data files (https://tcga-data.nci.nih.gov/docs/publications/luad_2014/). *TP53* was commonly mutated (46%). Mutations in *KRAS* (33%) were mutually exclusive with those in *EGFR* (14%). *BRAF* was also commonly mutated (10%), as were *PIK3CA* (7%), *MET* (7%) and the small GTPase gene, *RIT1* (2%). Mutations in tumour suppressor genes including *STK11* (17%), *KEAP1* (17%), *NF1* (11%), *RBI* (4%) and *CDKN2A* (4%) were observed. Mutations in chromatin modifying genes *SETD2* (9%), *ARID1A* (7%) and *SMARCA4* (6%) and the RNA splicing genes *RBM10* (8%) and *U2AF1* (3%) were also common. Recurrent mutations in the *MGA* gene (which encodes a Max-interacting protein on the MYC pathway¹⁹) occurred in 8% of samples. Loss-of-function (frameshift and nonsense) mutations in *MGA* were mutually exclusive with focal *MYC* amplification (Fisher's exact test $P = 0.04$), suggesting a hitherto unappreciated potential mechanism of MYC pathway activation. Coding single nucleotide variants and indel variants were verified by resequencing at a rate of 99% and 100%, respectively (Supplementary Fig. 3a, Supplementary Table 5). Tumour purity was not associated with the presence of false negatives identified in the validation data ($P = 0.31$; Supplementary Fig. 3b).

Past or present smoking associated with cytosine to adenine (C > A) nucleotide transversions as previously described both in individual genes and genome-wide^{12,13}. C > A nucleotide transversion fraction showed two peaks; this fraction correlated with total mutation count ($R^2 = 0.30$) and inversely correlated with cytosine to thymine (C > T) transition frequency ($R^2 = 0.75$) (Supplementary Fig. 4). We classified each sample (Supplementary Methods) into one of two groups named transversion-high (TH, $n = 269$), and transversion-low (TL, $n = 144$). The transversion-high group was strongly associated with past or present smoking ($P < 2.2 \times 10^{-16}$), consistent with previous reports¹³. The transversion-high and transversion-low patient cohorts harboured different gene mutations. Whereas *KRAS* mutations were significantly enriched in the transversion-high cohort ($P = 2.1 \times 10^{-13}$), *EGFR* mutations were significantly enriched in the transversion-low group ($P = 3.3 \times 10^{-6}$). *PIK3CA* and *RBI* mutations were likewise enriched in transversion-low tumours ($P < 0.05$). Additionally, the transversion-low tumours were specifically enriched for in-frame insertions in *EGFR* and *ERBB2* (ref. 5) and for frameshift indels in *RBI* (Fig. 1b). *RBI* is commonly mutated in small-cell lung carcinoma (SCLC). We found *RBI* mutations in transversion-low adenocarcinomas were enriched for frameshift indels versus single nucleotide substitutions compared to SCLC ($P < 0.05$)^{20,21} suggesting a mutational mechanism in transversion-low adenocarcinoma that is probably distinct from smoking in SCLC.

Gender is correlated with mutation patterns in lung adenocarcinoma²². Only a fraction of significantly mutated genes from the complete set reported in this study (Fig. 1a) were enriched in men or women (Fig. 1c). *EGFR* mutations were enriched in tumours from the female cohort ($P = 0.03$) whereas loss-of-function mutations within *RBM10*, an RNA-binding protein located on the X chromosome²³ were enriched in tumours from men ($P = 0.002$). When examining the transversion-high group, 16 out of 21 *RBM10* mutations were observed in males ($P = 0.003$, Fisher's exact test).

Somatic copy number alterations were very similar to those previously reported for lung adenocarcinoma²⁴ (Supplementary Fig. 5, Supplementary Table 6). Significant amplifications included *NKX2-1*, *TERT*,

MDM2, *KRAS*, *EGFR*, *MET*, *CCNE1*, *CCND1*, *TERC* and *MECOM* (Supplementary Table 6), as previously described²⁴, 8q24 near *MYC*, and a novel peak containing *CCND3* (Supplementary Table 6). The *CDKN2A* locus was the most significant deletion (Supplementary Table 6). Supplementary Table 7 summarizes molecular and clinical characteristics by sample. Low-pass whole-genome sequencing on a subset ($n = 93$) of the samples revealed an average of 36 gene–gene and gene–inter-gene

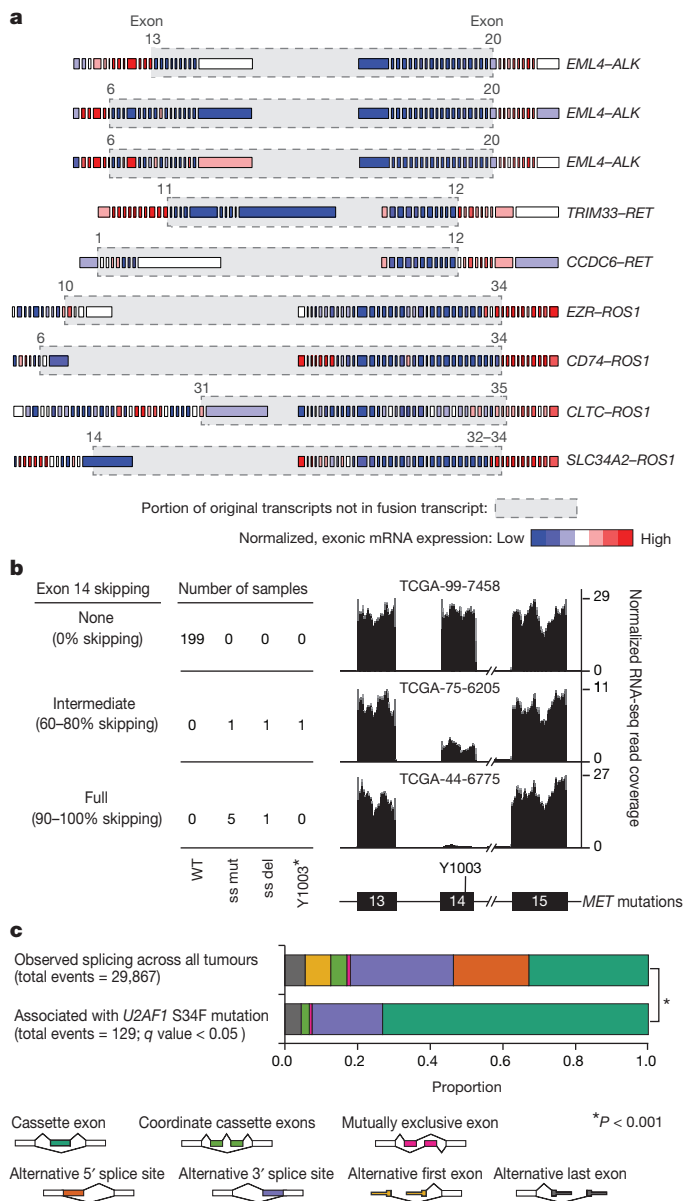


Figure 2 | Aberrant RNA transcripts in lung adenocarcinoma associated with somatic DNA translocation or mutation. **a**, Normalized exon level RNA expression across fusion gene partners. Grey boxes around genes mark the regions that are removed as a consequence of the fusion. Junction points of the fusion events are also listed in Supplementary Table 9. Exon numbers refer to reference transcripts listed in Supplementary Table 9. **b**, *MET* exon 14 skipping observed in the presence of exon 14 splice site mutation (ss mut), splice site deletion (ss del) or a Y1003* mutation. A total of 22 samples had insufficient coverage around exon 14 for quantification. The percentage skipping is (total expression minus exon 14 expression)/total expression. **c**, Significant differences in the frequency of 129 alternative splicing events in mRNA from tumours with *U2AF1* S34F compared to *U2AF1* WT tumours (q value < 0.05). Consistent with the function of *U2AF1* in 3' splice site recognition, most splicing differences involved cassette exon and alternative 3' splice site events (chi-squared test, $P < 0.001$).

rearrangements per tumour. Chromothripsis²⁵ occurred in six of the 93 samples (6%) (Supplementary Fig. 6, Supplementary Table 8). Low-pass whole genome sequencing-detected rearrangements appear in Supplementary Table 9.

Description of aberrant RNA transcripts

Gene fusions, splice site mutations or mutations in genes encoding splicing factors promote or sustain the malignant phenotype by generating aberrant RNA transcripts. Combining DNA with mRNA sequencing enabled us to catalogue aberrant RNA transcripts and, in many cases, to identify the DNA-encoded mechanism for the aberration. Seventy-five per cent of somatic mutations identified by WES were present in the RNA transcriptome when the locus in question was expressed (minimum 5×) (Supplementary Fig. 7a) similar to prior analyses¹⁵. Previously identified fusions involving *ALK* (3/230 cases), *ROS1* (4/230) and *RET* (2/230) (Fig. 2a, Supplementary Table 10), all occurred in transversion-low tumours ($P = 1.85 \times 10^{-4}$, Fisher's exact test).

MET activation can occur by exon 14 skipping, which results in a stabilized protein²⁶. Ten tumours had somatic *MET* DNA alterations with *MET* exon 14 skipping in RNA. In nine of these samples, a 5' or 3' splice site mutation or deletion was identified²⁷. *MET* exon 14 skipping was also found in the setting of a *MET* Y1003* stop codon mutation (Fig. 2b, Supplementary Fig. 8a). The codon affected by the Y1003* mutation is predicted to disrupt multiple splicing enhancer sequences, but the mechanism of skipping remains unknown in this case.

S34F mutations in *U2AF1* have recently been reported in lung adenocarcinoma¹² but their contribution to oncogenesis remains unknown. Eight samples harboured *U2AF1*^{S34F}. We identified 129 splicing events strongly associated with *U2AF1*^{S34F} mutation, consistent with the role of *U2AF1* in 3'-splice site selection²⁸. Cassette exons and alternative 3' splice sites were most commonly affected (Fig. 2c, Supplementary Table 11)²⁹. Among these events, alternative splicing of the *CTNNB1* proto-oncogene was strongly associated with *U2AF1* mutations (Supplementary Fig. 8b). Thus, concurrent analysis of DNA and RNA enabled delineation of both *cis* and *trans* mechanisms governing RNA processing in lung adenocarcinoma.

Candidate driver genes

The receptor tyrosine kinase (RTK)/RAS/RAF pathway is frequently mutated in lung adenocarcinoma. Striking therapeutic responses are often achieved when mutant pathway components are successfully inhibited. Sixty-two per cent (143/230) of tumours harboured known activating mutations in known driver oncogenes, as defined by others³⁰. Cancer-associated mutations in *KRAS* (32%, $n = 74$), *EGFR* (11%, $n = 26$) and *BRAF* (7%, $n = 16$) were common. Additional, previously uncharacterized *KRAS*, *EGFR* and *BRAF* mutations were observed, but were not classified as driver oncogenes for the purposes of our analyses (see Supplementary Fig. 9a for depiction of all mutations of known and unknown significance); explaining the differing mutation frequencies in each gene between this analysis and the overall mutational analysis described above. We also identified known activating *ERBB2* in-frame insertion and point mutations ($n = 5$)⁶, as well as mutations in *MAP2K1* ($n = 2$), *NRAS* and *HRAS* ($n = 1$ each). RNA sequencing revealed the aforementioned *MET* exon 14 skipping ($n = 10$) and fusions involving *ROS1* ($n = 4$), *ALK* ($n = 3$) and *RET* ($n = 2$). We considered these tumours collectively as oncogene-positive, as they harboured a known activating RTK/RAS/RAF pathway somatic event. DNA amplification events were not considered to be driver events before the comparisons described below.

We sought to nominate previously unrecognized genomic events that might activate this critical pathway in the 38% of samples without a RTK/RAS/RAF oncogene mutation. Tumour cellularity did not differ between oncogene-negative and oncogene-positive samples (Supplementary Fig. 9b). Analysis of copy number alterations using GISTIC³¹ identified unique focal *ERBB2* and *MET* amplifications in the oncogene-negative subset (Fig. 3a, Supplementary Table 6); amplifications in other wild-type proto-oncogenes, including *KRAS* and *EGFR*, were not significantly different between the two groups.

We next analysed WES data independently in the oncogene-negative and oncogene-positive subsets. We found that *TP53*, *KEAP1*, *NF1* and *RIT1* mutations were significantly enriched in oncogene-negative tumours ($P < 0.01$; Fig. 3b, Supplementary Table 12). *NF1* mutations have previously been reported in lung adenocarcinoma¹¹, but this is the first study, to our knowledge, capable of identifying all classes of loss-of-function

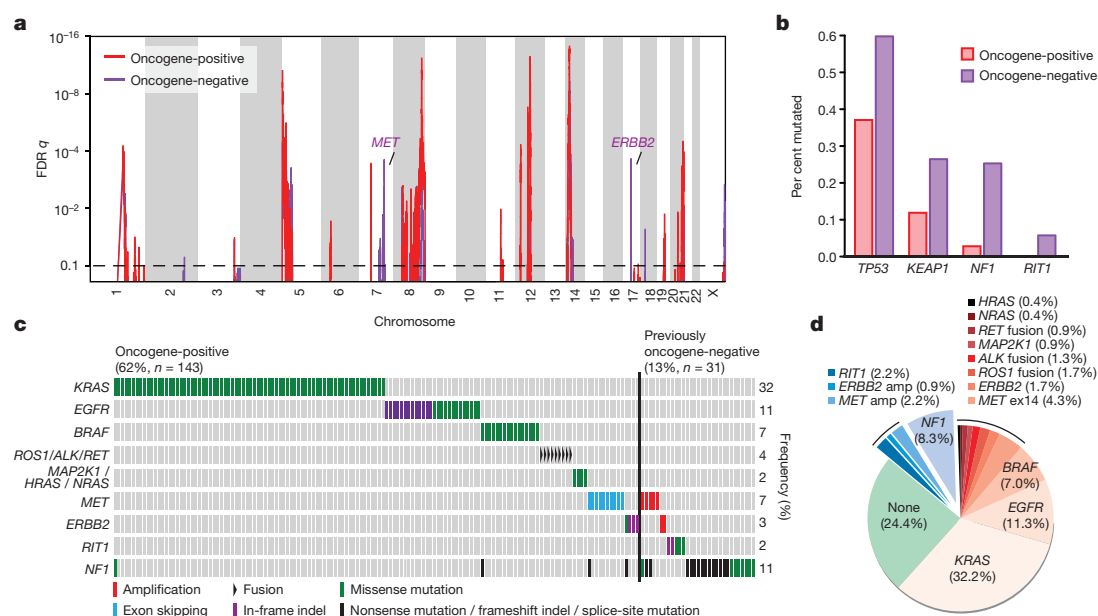
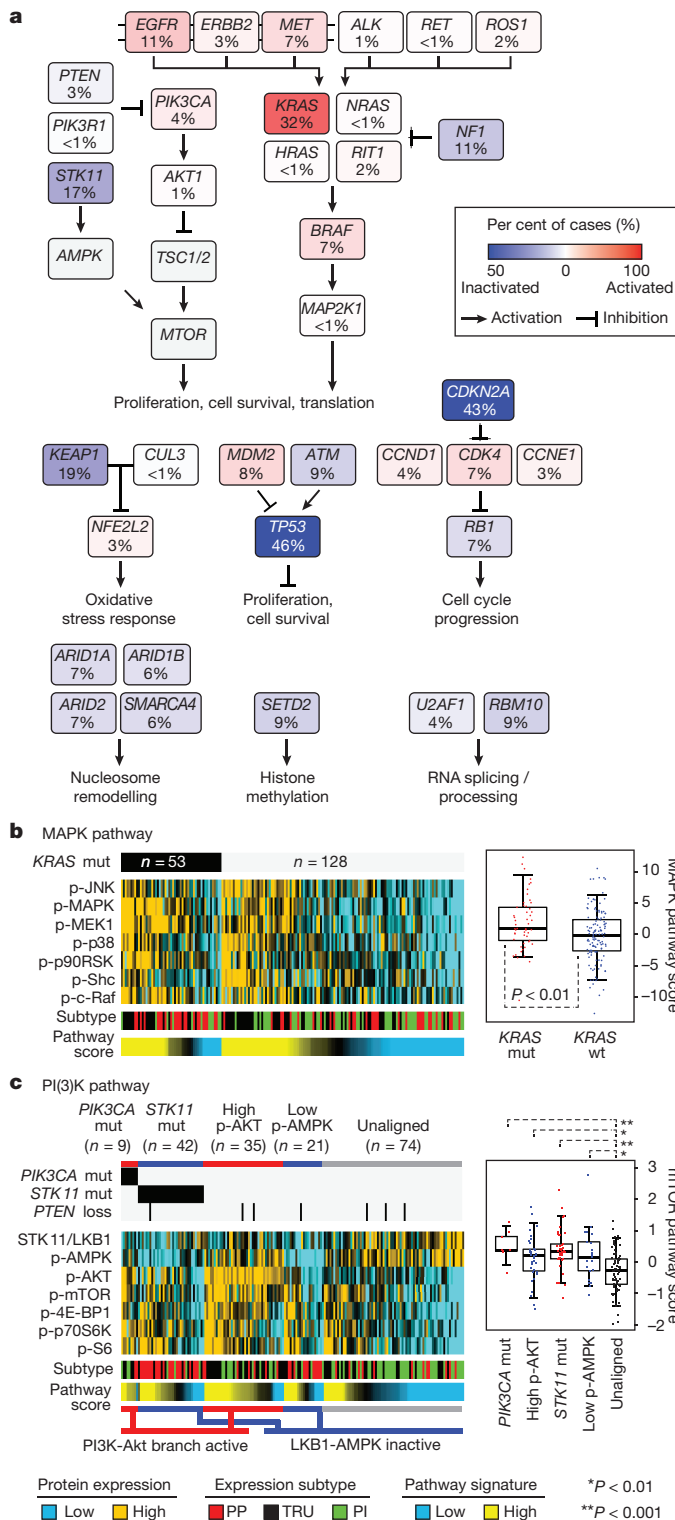


Figure 3 | Identification of novel candidate driver genes. **a**, GISTIC analysis of focal amplifications in oncogene-negative ($n = 87$) and oncogene-positive ($n = 143$) TCGA samples identifies focal gains of *MET* and *ERBB2* that are specific to the oncogene-negative set (purple). **b**, *TP53*, *KEAP1*, *NF1* and *RIT1* mutations are significantly enriched in samples otherwise lacking oncogene mutations (adjusted $P < 0.05$ by Fisher's exact test). **c**, Co-mutation plot of variants of known significance within the RTK/RAS/RAF pathway in lung

adenocarcinoma. Not shown are the 63 tumours lacking an identifiable driver lesion. Only canonical driver events, as defined in Supplementary Fig. 9, and proposed driver events, are shown; hence not every alteration found is displayed. **d**, New candidate driver oncogenes (blue: 13% of cases) and known somatically activated drivers events (red: 63%) that activate the RTK/RAS/RAF pathway can be found in the majority of the 230 lung adenocarcinomas.

NF1 defects and to statistically demonstrate that *NF1* mutations, as well as *KEAP1* and *TP53* mutations are enriched in the oncogene-negative subset of lung adenocarcinomas (Fig. 3c). All *RIT1* mutations occurred in the oncogene-negative subset and clustered around residue Q79 (homologous to Q61 in the switch II region of *RAS* genes). These mutations transform NIH3T3 cells and activate MAPK and PI(3)K signalling³², supporting a driver role for mutant *RIT1* in 2% of lung adenocarcinomas. This analysis increases the rate at which putative somatic lung adenocarcinoma driver events can be identified within the RTK/RAS/RAF pathway to 76% (Fig. 3d).



Recurrent alterations in key pathways

Recurrent aberrations in multiple key pathways and processes characterize lung adenocarcinoma (Fig. 4a). Among these were RTK/RAS/RAF pathway activation (76% of cases), PI(3)K-mTOR pathway activation (25%), p53 pathway alteration (63%), alteration of cell cycle regulators (64%, Supplementary Fig. 10), alteration of oxidative stress pathways (22%, Supplementary Fig. 11), and mutation of various chromatin and RNA splicing factors (49%).

We then examined the phenotypic sequelae of some key genomic events in the tumours in which they occurred. Reverse-phase protein arrays provided proteomic and phosphoproteomic phenotypic evidence of pathway activity. Antibodies on this platform are listed in Supplementary Table 13. This analysis suggested that DNA sequencing did not identify all samples with phosphoprotein evidence of activation of a given signalling pathway. For example, whereas *KRAS*-mutant lung adenocarcinomas had higher levels of phosphorylated MAPK than *KRAS* wild-type tumours had on average, many *KRAS* wild-type tumours displayed significant MAPK pathway activation (Fig. 4b, Supplementary Fig. 10). The multiple mechanisms by which lung adenocarcinomas achieve MAPK activation suggest additional, still undetected RTK/RAS/RAF pathway alterations. Similarly, we found significant activation of mTOR and its effectors (p70S6kinase, S6, 4E-BP1) in a substantial fraction of the tumours (Fig. 4c). Analysis of mutations in *PIK3CA* and *STK11*, *STK11* protein levels, and AMPK and AKT phosphorylation³³ led to the identification of three major mTOR patterns in lung adenocarcinoma: (1) tumours with minimal or basal mTOR pathway activation, (2) tumours showing higher mTOR activity accompanied by either *STK11*-inactivating mutation or combined low *STK11* expression and low AMPK activation and (3) tumours showing high mTOR activity accompanied by either phosphorylated AKT activation, *PIK3CA* mutation, or both. As with MAPK, many tumours lack an obvious underlying genomic alteration to explain their apparent mTOR activation.

Molecular subtypes of lung adenocarcinoma

Broad transcriptional and epigenetic profiling can reveal downstream consequences of driver mutations, provide clinically relevant classification and offer insight into tumours lacking clear drivers. Prior unsupervised analyses of lung adenocarcinoma gene expression have used varying nomenclature for transcriptional subtypes of the disease^{34–37}. To coordinate naming of the transcriptional subtypes with the histopathological³⁸, anatomic and mutational classifications of lung adenocarcinoma, we propose an updated nomenclature: the terminal respiratory unit (TRU, formerly bronchioid), the proximal-inflammatory (PI, formerly squamoid), and the proximal-proliferative (PP, formerly magnoid)³⁹ transcriptional subtypes (Fig. 5a). Previously reported associations of expression signatures with pathways and clinical outcomes^{34,36,39} were observed (Supplementary Fig. 7b) and integration with multi-analyte data revealed statistically significant genomic alterations associated with these transcriptional subtypes. The PP subtype was enriched for mutation of *KRAS*, along with inactivation of the *STK11* tumour suppressor gene by chromosomal loss, inactivating mutation, and reduced gene expression. In contrast, the PI subtype was characterized by solid histopathology and

Figure 4 | Pathway alterations in lung adenocarcinoma. **a**, Somatic alterations involving key pathway components for RTK signalling, mTOR signalling, oxidative stress response, proliferation and cell cycle progression, nucleosome remodelling, histone methylation, and RNA splicing/processing. **b**, **c**, Proteomic analysis by RPPA ($n = 181$) P values by two-sided t -test. Box plots represent 5%, 25%, 75%, median, and 95%. PP, proximal proliferative; TRU, terminal respiratory unit; PI, proximal inflammatory. **c**, mTOR signalling may be activated, by either Akt (for example, via PI(3)K) or inactivation of AMPK (for example, via *STK11* loss). Tumours were separated into three main groups: those with PI(3)K-AKT activation, through either *PIK3CA* activating mutation or unknown mechanism (high p-AKT); those with LKB1-AMPK inactivation, through either *STK11* mutation or unknown mechanism with low levels of LKB1 and p-AMPK; and those showing none of the above features.

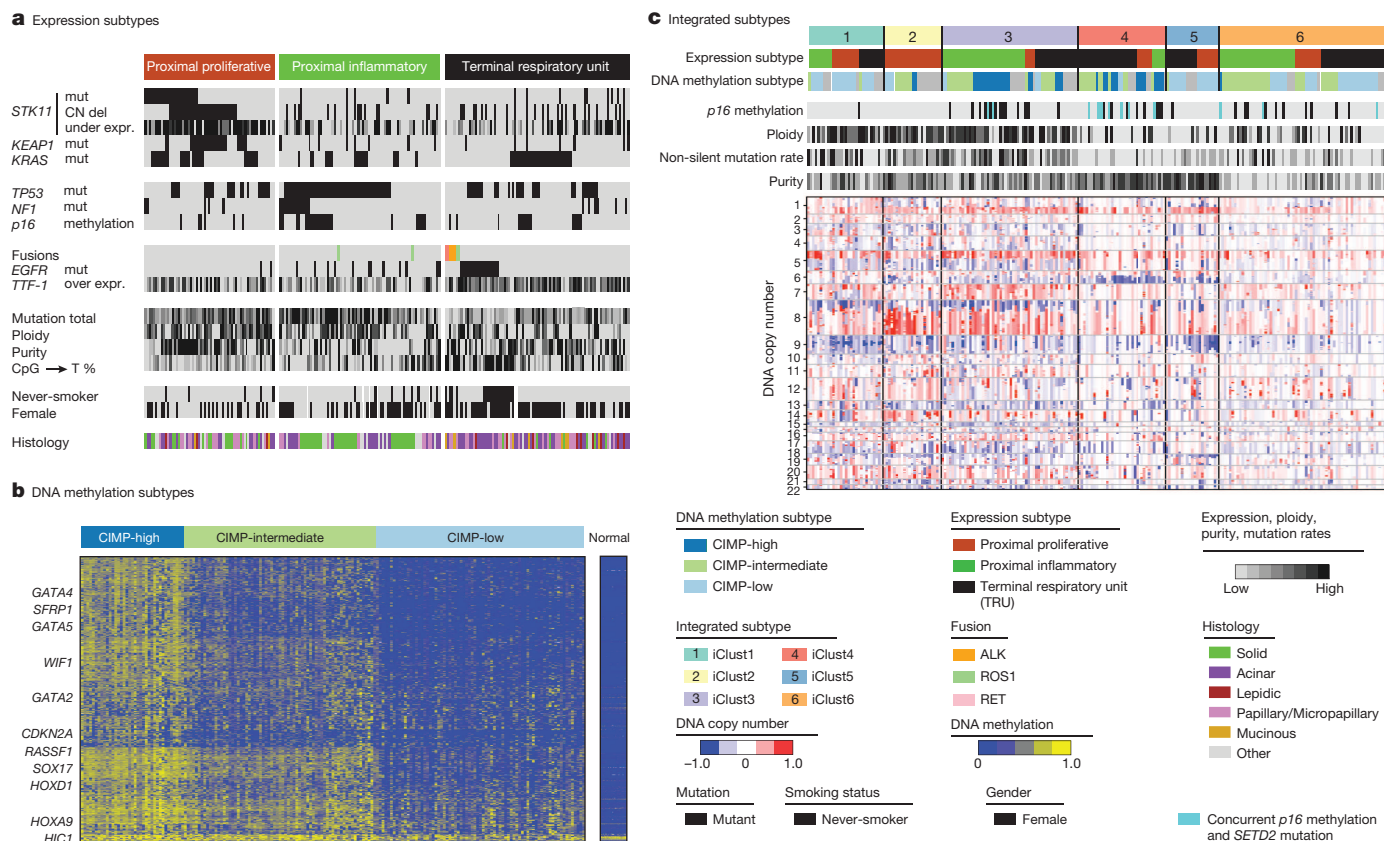


Figure 5 | Integrative analysis. a–c, Integrating unsupervised analyses of 230 lung adenocarcinomas reveals significant interactions between molecular subtypes. Tumours are displayed as columns, grouped by mRNA expression subtypes (a), DNA methylation subtypes (b), and integrated subtypes by

iCluster analysis (c). All displayed features are significantly associated with subtypes depicted. The CIMP phenotype is defined by the most variable CpG island and promoter probes.

co-mutation of *NF1* and *TP53*. Finally, the TRU subtype harboured the majority of the *EGFR*-mutated tumours as well as the kinase fusion expressing tumours. TRU subtype membership was prognostically favourable, as seen previously³⁴ (Supplementary Fig. 7c). Finally, the subtypes exhibited different mutation rates, transition frequencies, genomic ploidy profiles, patterns of large-scale aberration, and differed in their association with smoking history (Fig. 5a). Unsupervised clustering of miRNA sequencing-derived or reverse phase protein array (RPPA)-derived data also revealed significant heterogeneity, partially overlapping with the mRNA-based subtypes, as demonstrated in Supplementary Figs 12 and 13.

Mutations in chromatin-modifying genes (for example, *SMARCA4*, *ARID1A* and *SETD2*) suggest a major role for chromatin maintenance in lung adenocarcinoma. To examine chromatin states in an unbiased manner, we selected the most variable DNA methylation-specific probes in CpG island promoter regions and clustered them by methylation intensity (Supplementary Table 14). This analysis divided samples into two distinct subsets: a significantly altered CpG island methylator phenotype-high (CIMP-H(igh)) cluster and a more normal-like CIMP-L(ow) group, with a third set of samples occupying an intermediate level of methylation at CIMP sites (Fig. 5b). Our results confirm a prior report⁴⁰ and provide additional insights into this epigenetic program. CIMP-H tumours often showed DNA hypermethylation of several key genes: *CDKN2A*, *GATA4*, *GATA5*, *HIC1*, *HOXA9*, *HOXD13*, *RASSF1*, *SFRP1*, *SOX17* and *WIF1* among others (Supplementary Fig. 14). WNT pathway genes are significantly over-represented in this list (P value = 0.0015) suggesting that this is a key pathway with an important driving role within this subtype. *MYC* overexpression was significantly associated with the CIMP-H phenotype as well (P = 0.003).

Although we did not find significant correlations between global DNA methylation patterns and individual mutations in chromatin remodelling genes, there was an intriguing association between *SETD2* mutation

and *CDKN2A* methylation. Tumours with low *CDKN2A* expression due to methylation (rather than due to mutation or deletion) had lower ploidy, fewer overall mutations (Fig. 5c) and were significantly enriched for *SETD2* mutation, suggesting an important role for this chromatin-modifying gene in the development of certain tumours.

Integrative clustering⁴¹ of copy number, DNA methylation and mRNA expression data found six clusters (Fig. 5c). Tumour ploidy and mutation rate are higher in clusters 1–3 than in clusters 4–6. Clusters 1–3 frequently harbour *TP53* mutations and are enriched for the two proximal transcriptional subtypes. Fisher's combined probability tests revealed significant copy number associated gene expression changes on 3q in cluster one, 8q in cluster two, and chromosome 7 and 15q in cluster three (Supplementary Fig. 15). The low ploidy and low mutation rate clusters four and five contain many TRU samples, whereas tumours in cluster 6 have comparatively lower tumour cellularity, and few other distinguishing molecular features. Significant copy number-associated gene expression changes are observed on 6q in cluster four and 19p in cluster five. The CIMP-H tumours divided into a high ploidy, high mutation rate, proximal-inflammatory CIMP-H group (cluster 3) and a low ploidy, low mutation rate, TRU-associated CIMP-H group (cluster 4), suggesting that the CIMP phenotype in lung adenocarcinoma can occur in markedly different genomic and transcriptional contexts. Furthermore, cluster four is enriched for *CDKN2A* methylation and *SETD2* mutations, suggesting an interaction between somatic mutation of *SETD2* and deregulated chromatin maintenance in this subtype. Finally, cluster membership was significantly associated with mutations in *TP53*, *EGFR* and *STK11* (Supplementary Fig. 15, Supplementary Table 6).

Conclusions

We assessed the mutation profiles, structural rearrangements, copy number alterations, DNA methylation, mRNA, miRNA and protein expression

of 230 lung adenocarcinomas. In recent years, the treatment of lung adenocarcinoma has been advanced by the development of multiple therapies targeted against alterations in the RTK/RAS/RAF pathway. We nominate amplifications in *MET* and *ERBB2* as well as mutations of *NF1* and *RIT1* as driver events specifically in otherwise oncogene-negative lung adenocarcinomas. This analysis increases the fraction of lung adenocarcinoma cases with somatic evidence of RTK/RAS/RAF activation from 62% to 76%. While all lung adenocarcinomas may activate this pathway by some mechanism, only a subset show tonic pathway activation at the protein level, suggesting both diversity between tumours with seemingly similar activating events and as yet undescribed mechanisms of pathway activation. Therefore, the current study expands the range of possible targetable alterations within the RTK/RAS/RAF pathway in general and suggests increased implementation of *MET* and *ERBB2/HER2* inhibitors in particular. Our discovery of inactivating mutations of *MGA* further underscores the importance of the *MYC* pathway in lung adenocarcinoma.

This study further implicates both chromatin modifications and splicing alterations in lung adenocarcinoma through the integration of DNA, transcriptome and methylome analysis. We identified alternative splicing due to both splicing factor mutations in *trans* and mutation of splice sites in *cis*, the latter leading to activation of the *MET* gene by exon 14 skipping. Cluster analysis separated tumours based on single-gene driver events as well as large-scale aberrations, emphasizing lung adenocarcinoma's molecular heterogeneity and combinatorial alterations, including the identification of coincident *SETD2* mutations and *CDKN2A* methylation in a subset of CIMP-H tumours, providing evidence of a somatic event associated with a genome-wide methylation phenotype. These studies provide new knowledge by illuminating modes of genomic alteration, highlighting previously unappreciated altered genes, and enabling further refinement in sub-classification for the improved personalization of treatment for this deadly disease.

METHODS SUMMARY

All specimens were obtained from patients with appropriate consent from the relevant institutional review board. DNA and RNA were collected from samples using the Allprep kit (Qiagen). We used standard approaches for capture and sequencing of exomes from tumour DNA and normal DNA¹⁵ and whole-genome shotgun sequencing. Significantly mutated genes were identified by comparing them with expectation models based on the exact measured rates of specific sequence lesions⁴². GISTIC analysis of the circular-binary-segmented Affymetrix SNP 6.0 copy number data was used to identify recurrent amplification and deletion peaks³¹. Consensus clustering approaches were used to analyse mRNA, miRNA and methylation subtypes using previous approaches¹⁵. The publication web page is (https://tcga-data.nci.nih.gov/docs/publications/luad_2014/). Sequence files are in CGHub (<https://cghub.ucsc.edu/>).

Received 11 June 2013; accepted 22 April 2014.

Published online 9 July 2014.

- Paez, J. G. *et al.* *EGFR* mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* **304**, 1497–1500 (2004).
- Kwak, E. L. *et al.* Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N. Engl. J. Med.* **363**, 1693–1703 (2010).
- Bergethon, K. *et al.* ROS1 rearrangements define a unique molecular class of lung cancers. *J. Clin. Oncol.* **30**, 863–870 (2012).
- Drilon, A. *et al.* Response to cabozantinib in patients with RET fusion-positive lung adenocarcinomas. *Cancer Discov.* **3**, 630–635 (2013).
- Stephens, P. *et al.* Lung cancer: intragenic *ERBB2* kinase mutations in tumours. *Nature* **431**, 525–526 (2004).
- Takahashi, T. *et al.* p53: a frequent target for genetic abnormalities in lung cancer. *Science* **246**, 491–494 (1989).
- Sanchez-Cespedes, M. *et al.* Inactivation of LKB1/STK11 is a common event in adenocarcinomas of the lung. *Cancer Res.* **62**, 3659–3662 (2002).
- Shapiro, G. I. *et al.* Reciprocal Rb inactivation and p16^{INK4} expression in primary lung cancers and cell lines. *Cancer Res.* **55**, 505–509 (1995).
- Singh, A. *et al.* Dysfunctional KEAP1–NRF2 interaction in non-small-cell lung cancer. *PLoS Med.* **3**, e420 (2006).
- Medina, P. P. *et al.* Frequent *BRG1/SMARCA4*-inactivating mutations in human lung cancer cell lines. *Hum. Mutat.* **29**, 617–622 (2008).
- Ding, L. *et al.* Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* **455**, 1069–1075 (2008).
- Imielinski, M. *et al.* Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* **150**, 1107–1120 (2012).

- Govindan, R. *et al.* Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell* **150**, 1121–1134 (2012).
- Travis, W. D., Brambilla, E. & Riely, G. J. New pathologic classification of lung cancer: relevance for clinical practice and clinical trials. *J. Clin. Oncol.* **31**, 992–1001 (2013).
- The Cancer Genome Atlas Research Network Comprehensive genomic characterization of squamous cell lung cancers. *Nature* **489**, 519–525 (2012).
- Carter, S. L. *et al.* Absolute quantification of somatic DNA alterations in human cancer. *Nature Biotechnol.* **30**, 413–421 (2012).
- Cibulskis, K. *et al.* Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nature Biotechnol.* **31**, 213–219 (2013).
- Lawrence, M. S. *et al.* Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* **505**, 495–501 (2014).
- Hurlin, P. J., Steingrimsson, E., Copeland, N. G., Jenkins, N. A. & Eisenman, R. N. Mga, a dual-specificity transcription factor that interacts with Max and contains a T-domain DNA-binding motif. *EMBO J.* **18**, 7019–7028 (1999).
- Peifer, M. *et al.* Integrative genome analyses identify key somatic driver mutations of small-cell lung cancer. *Nature Genet.* **44**, 1104–1110 (2012).
- Rudin, C. M. *et al.* Comprehensive genomic analysis identifies *SOX2* as a frequently amplified gene in small-cell lung cancer. *Nature Genet.* **44**, 1111–1116 (2012).
- Tokumo, M. *et al.* The relationship between epidermal growth factor receptor mutations and clinicopathologic features in non-small cell lung cancers. *Clin. Cancer Res.* **11**, 1167–1173 (2005).
- Coleman, M. P. *et al.* A novel gene, DXS8237E, lies within 20 kb upstream of UBE1 in Xp11.23 and has a different X inactivation status. *Genomics* **31**, 135–138 (1996).
- Weir, B. A. *et al.* Characterizing the cancer genome in lung adenocarcinoma. *Nature* **450**, 893–898 (2007).
- Stephens, P. J. *et al.* Massive genomic rearrangement acquired in a single catastrophic event during cancer development. *Cell* **144**, 27–40 (2011).
- Kong-Beltran, M. *et al.* Somatic mutations lead to an oncogenic deletion of Met in lung cancer. *Cancer Res.* **66**, 283–289 (2006).
- Seo, J. S. *et al.* The transcriptional landscape and mutational profile of lung adenocarcinoma. *Genome Res.* **22**, 2109–2119 (2012).
- Wu, S., Romfo, C. M., Nilsen, T. W. & Green, M. R. Functional recognition of the 3' splice site AG by the splicing factor U2AF³⁵. *Nature* **402**, 832–835 (1999).
- Brooks, A. N. *et al.* A pan-cancer analysis of transcriptome changes associated with somatic mutations in *U2AF1* reveals commonly altered splicing events. *PLoS ONE* **9**, e87361 (2014).
- Pao, W. & Hutchinson, K. E. Chipping away at the lung cancer genome. *Nature Med.* **18**, 349–351 (2012).
- Beroukhim, R. *et al.* Assessing the significance of chromosomal aberrations in cancer: methodology and application to glioma. *Proc. Natl Acad. Sci. USA* **104**, 20007–20012 (2007).
- Berger, A. H. *et al.* Oncogenic *RIT1* mutations in lung adenocarcinoma. *Oncogene* <http://dx.doi.org/10.1038/ncr.2013.581> (2014).
- Creighton, C. J. *et al.* Proteomic and transcriptomic profiling reveals a link between the PI3K pathway and lower estrogen-receptor (ER) levels and activity in ER⁺ breast cancer. *Breast Cancer Res.* **12**, R40 (2010).
- Wilkerson, M. D. *et al.* Differential pathogenesis of lung adenocarcinoma subtypes involving sequence mutations, copy number, chromosomal instability, and methylation. *PLoS ONE* **7**, e36530 (2012).
- Beer, D. G. *et al.* Gene-expression profiles predict survival of patients with lung adenocarcinoma. *Nature Med.* **8**, 816–824 (2002).
- Hayes, D. N. *et al.* Gene expression profiling reveals reproducible human lung adenocarcinoma subtypes in multiple independent patient cohorts. *J. Clin. Oncol.* **24**, 5079–5090 (2006).
- Bhattacharjee, A. *et al.* Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses. *Proc. Natl Acad. Sci. USA* **98**, 13790–13795 (2001).
- Travis, W. D. *et al.* International association for the study of lung cancer/American Thoracic Society/European Respiratory Society international multidisciplinary classification of lung adenocarcinoma. *J. Thoracic Oncol.* **6**, 244–285 (2011).
- Yatabe, Y., Mitsudomi, T. & Takahashi, T. TTF-1 expression in pulmonary adenocarcinomas. *Am. J. Surg. Pathol.* **26**, 767–773 (2002).
- Shinjo, K. *et al.* Integrated analysis of genetic and epigenetic alterations reveals CpG island methylator phenotype associated with distinct clinical characters of lung adenocarcinoma. *Carcinogenesis* **33**, 1277–1285 (2012).
- Mo, Q. *et al.* Pattern discovery and cancer gene identification in integrated cancer genomic data. *Proc. Natl Acad. Sci. USA* **110**, 4245–4250 (2013).
- Lawrence, M. S. *et al.* Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature* **499**, 214–218 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements This study was supported by NIH grants: U24 CA126561, U24 CA126551, U24 CA126554, U24 CA126543, U24 CA126546, U24 CA137153, U24 CA126563, U24 CA126544, U24 CA143845, U24 CA143858, U24 CA144025, U24 CA143882, U24 CA143866, U24 CA143867, U24 CA143848, U24 CA143840, U24 CA143835, U24 CA143799, U24 CA143883, U24 CA143843, U54 HG003067, U54 HG003079 and U54 HG003273. We thank K. Guebert and L. Gaffney for assistance and C. Gunter for review.

Author Contributions The Cancer Genome Atlas Research Network contributed collectively to this study. Biospecimens were provided by the tissue source sites and processed by the biospecimen core resource. Data generation and analyses were performed by the genome sequencing centres, cancer genome characterization centres and genome data analysis centres. All data were released through the data coordinating centre. The National Cancer Institute and National Human Genome Research Institute project teams coordinated project activities. We also acknowledge the following TCGA investigators who made substantial contributions to the project: E. A. Collisson (manuscript coordinator); J. D. Campbell, J. Chmielecki, (analysis coordinators); C. Sougnez (data coordinator); J. D. Campbell, M. Rosenberg, W. Lee, J. Chmielecki, M. Ladanyi, and G. Getz (DNA sequence analysis); M. D. Wilkerson, A. N. Brooks, and D. N. Hayes (mRNA sequence analysis); L. Danilova and L. Cope (DNA methylation analysis); A. D. Cherniack (copy number analysis); M. D. Wilkerson and A. Hadjipanayis (translocations); N. Schultz, W. Lee, E. A. Collisson, A. H. Berger, J. Chmielecki, C. J. Creighton, L. A. Byers and M. Ladanyi (pathway analysis); A. Chu and A. G. Robertson (miRNA sequence analysis); W. Travis and D. A. Wagle (pathology and clinical expertise); L. A. Byers and G. B. Mills (reverse phase protein arrays); S. B. Baylin, R. Govindan and M. Meyerson (project chairs).

Author Information The primary and processed data used to generate the analyses presented here can be downloaded by registered users from The Cancer Genome Atlas at (<https://tcga-data.nci.nih.gov/tcga/tcgaDownload.jsp>). All of the primary sequence files are deposited in cGHub and all other data are deposited at the Data Coordinating Center (DCC) for public access (<http://cancergenome.nih.gov/>), (<https://cghub.ucsc.edu/>) and (https://tcga-data.nci.nih.gov/docs/publications/luad_2014/). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.M. (matthew_meyerson@dfci.harvard.edu).



This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons licence, users will need to obtain permission from the licence holder to reproduce the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

The Cancer Genome Atlas Research Network

Disease analysis working group Eric A. Collisson¹, Joshua D. Campbell², Angela N. Brooks^{2,3}, Alice H. Berger², William Lee⁴, Juliann Chmielecki², David G. Beer², Leslie Cope⁶, Chad J. Creighton⁷, Ludmila Danilova⁶, Li Ding⁸, Gad Getz^{2,9,10}, Peter S. Hammerman², D. Neil Hayes¹¹, Bryan Hernandez², James G. Herman⁶, John V. Heymach¹², Igor Jurisica¹³, Raju Kucherlapati⁹, David Kwiatkowski¹⁴, Marc Ladanyi⁴, Gordon Robertson¹⁵, Nikolaus Schultz⁴, Ronglai Shen⁴, Rileen Sinha¹², Carrie Sougnez², Ming-Sound Tsao¹³, William D. Travis⁴, John N. Weinstein¹², Dennis A. Wagle¹⁶, Matthew D. Wilkerson¹¹, Andy Chu¹⁵, Andrew D. Cherniack², Angela Hadjipanayis⁹, Mara Rosenberg², Daniel J. Weisenberger¹⁷, Peter W. Laird¹⁷, Amie Radenbaugh¹⁸, Singer Ma¹⁸, Joshua M. Stuart¹⁸, Lauren Averett Byers¹², Stephen B. Baylin⁶, Ramaswamy Govindan⁸, Matthew Meyerson^{2,3}

Genome sequencing centres: The Eli & Edythe L. Broad Institute Mara Rosenberg², Stacey B. Gabriel², Kristian Cibulskis², Carrie Sougnez², Jaegil Kim², Chip Stewart², Lee Lichtenstein², Eric S. Lander^{2,19}, Michael S. Lawrence², Getz^{2,9,10}, **Washington University in St. Louis** Cyriac Kandoth⁸, Robert Fulton⁸, Lucinda L. Fulton⁸, Michael D. McLellan⁸, Richard K. Wilson⁸, Kai Ye⁸, Catrina C. Fronick⁸, Christopher A. Maher⁸, Christopher A. Miller⁸, Michael C. Wendt⁸, Christopher Cabanski⁸, Li Ding⁸, Elaine Mardis⁸, Ramaswamy Govindan⁸, **Baylor College of Medicine** Chad J. Creighton⁷, David Wheeler⁷

Genome characterization centres: Canada's Michael Smith Genome Sciences Centre, British Columbia Cancer Agency Miruna Balasundaram¹⁵, Yaron S. N. Butterfield¹⁵, Rebecca Carlsen¹⁵, Andy Chu¹⁵, Eric Chuah¹⁵, Noreen Dhalla¹⁵, Ranabir Guin¹⁵, Carrie Hirst¹⁵, Darlene Lee¹⁵, Haiyan I. Li¹⁵, Michael Mayo¹⁵, Richard A. Moore¹⁵, Andrew J. Mungall¹⁵, Jacqueline E. Schein¹⁵, Payal Sipahimalani¹⁵, Angela Tam¹⁵, Richard Varhol¹⁵, A. Gordon Robertson¹⁵, Natassja Wye¹⁵, Nina Thiessen¹⁵, Robert A. Holt¹², Steven J. M. Jones¹⁵, Marco A. Marra¹⁵, **The Eli & Edythe L. Broad Institute** Joshua D. Campbell², Angela N. Brooks^{2,3}, Juliann Chmielecki², Marc Imielinski^{2,9,10}, Robert C. Onofri², Eran Hodis³, Travis Zack², Carrie Sougnez², Elena Helman², Chandra Sekhar Pedamallu², Jill Mesirov², Andrew D. Cherniack², Gordon Saksena², Steven E. Schumacher², Scott L. Carter², Bryan Hernandez², Levi Garraway^{2,3,9}, Rameen Beroukhi^{2,3,9}, Stacey B. Gabriel², Gad Getz^{2,9,10}, Matthew Meyerson^{2,3,9}, **Harvard Medical School/Brigham & Women's Hospital/MD Anderson Cancer Center** Angela Hadjipanayis^{9,14}, Semin Lee^{9,14}, Harshad S. Mahadeshwar¹², Angeliki Pantazi^{9,14}, Alexei Protopopov¹², Xiaojia Ren⁹, Sahil Seth¹², Xingzhi Song¹², Jiabin Tang¹², Lixing Yang⁹, Jianhua Zhang¹², Peng-Chieh Chen⁹, Michael Parfenov^{9,14}, Andrew Wei Xu^{9,14}, Netty Santoso^{9,14}, Lynda Chin¹², Peter J. Park^{9,14} & Raju Kucherlapati^{9,14}, **University of North Carolina, Chapel Hill** Katherine A. Hoadley¹¹, J. Todd Auman¹¹, Shaowu Meng¹¹, Yan Shi¹¹, Elizabeth Buda¹¹, Scot Waring¹¹, Umadevi Veluvolu¹¹, Donghui Tan¹¹, Piotr A. Mieczkowski¹¹, Corbin D. Jones¹¹, Janae V. Simons¹¹, Matthew G. Soloway¹¹, Tom Bodenheimer¹¹, Stuart R. Jefferys¹¹, Jeffrey Roach¹¹, Alan P. Hoyle¹¹, Junyuan Wu¹¹, Saianand Balu¹¹, Darshan Singh¹¹, Jan F.

Prins¹¹, J.S. Marron¹¹, Joel S. Parker¹¹, D. Neil Hayes¹¹, Charles M. Perou¹¹; **University of Kentucky** Jinze Liu²⁰; **The USC/JHU Epigenome Characterization Center** Leslie Cope⁶, Ludmila Danilova⁶, Daniel J. Weisenberger¹⁷, Dennis T. Maglinte¹⁷, Philip H. Lai¹⁷, Moiz S. Bootwalla¹⁷, David J. Van Den Berg¹⁷, Timothy Triche Jr¹⁷, Stephen B. Baylin⁶, Peter W. Laird¹⁷

Genome data analysis centres: The Eli & Edythe L. Broad Institute Mara Rosenberg², Lynda Chin¹², Jianhua Zhang¹², Juok Cho², Daniel DiCara², David Heiman², Pei Lin², William Mallard², Douglas Voet², Hailei Zhang², Lihua Zou², Michael S. Noble², Michael S. Lawrence², Gordon Saksena², Nils Gehlenborg², Helga Thorvaldsdottir², Jill Mesirov², Marc-Danie Nazaire², Jim Robinson², Gad Getz^{2,9,10}, **Memorial Sloan-Kettering Cancer Center** William Lee⁴, B. Arman Aksoy⁴, Giovanni Ciriello⁴, Barry S. Taylor⁴, Gideon Dresdner⁴, Jianjiong Gao⁴, Benjamin Gross⁴, Venkatraman E. Seshan⁴, Marc Ladanyi⁴, Boris Reva⁴, Rileen Sinha⁴, S. Onur Sumer⁴, Nils Wenhöld⁴, Nikolaus Schultz⁴, Ronglai Shen⁴, Chris Sander⁴; **University of California, Santa Cruz/Buck Institute** Sam Ng¹⁸, Singer Ma¹⁸, Jingchun Zhu¹⁸, Amie Radenbaugh¹⁸, Joshua M. Stuart¹⁸, Christopher C. Benz²¹, Christina Yau²¹ & David Haussler^{18,22}, **Oregon Health & Sciences University** Paul T. Spellman²³; **University of North Carolina, Chapel Hill** Matthew D. Wilkerson¹¹, Joel S. Parker¹¹, Katherine A. Hoadley¹¹, Patrick K. Kimes¹¹, D. Neil Hayes¹¹, Charles M. Perou¹¹, **The University of Texas MD Anderson Cancer Center** Bradley M. Broom¹², Jing Wang¹², Yiling Lu¹², Patrick Kwok Shing Ng¹², Lixia Diao¹², Lauren Averett Byers¹², Wenbin Liu¹², John V. Heymach¹², Christopher I. Amos¹², John N. Weinstein¹², Rehan Akbani¹², Gordon B. Mills¹²

Biospecimen core resource: International Genomics Consortium Erin Curley²⁴, Joseph Paulauskas²⁴, Kevin Lau²⁴, Scott Morris²⁴, Troy Shelton²⁴, David Mallory²⁴, Johanna Gardner²⁴, Robert Penny²⁴

Tissue source sites: Analytical Biological Service, Inc. Charles Saller²⁵, Katherine Tarvin²⁵; **Brigham & Women's Hospital** William G. Richards¹⁴; **University of Alabama at Birmingham** Robert Cerfolio²⁶, Ayesha Bryant²⁶; **Cleveland Clinic**: Daniel P. Raymond²⁷, Nathan A. Pennell²⁷, Carol Farver²⁷; **Christiana Care** Christine Czerwinski²⁸, Lori Huelsenbeck-Dill²⁸, Mary Iacocca²⁸, Nicholas Petrelli²⁸, Brenda Rabeno²⁸, Jennifer Brown²⁸, Thomas Bauer²⁸; **Cureline** Oleg Dolzhanskiy²⁹, Olga Potapova²⁹, Daniil Rotin²⁹, Olga Voronina²⁹, Elena Nemirovich-Danchenko²⁹, Konstantin V. Fedosenko²⁹; **Emory University** Anthony Gai³⁰, Madhusmita Behera³⁰, Suresh S. Ramalingam³⁰, Gabriel Sica³⁰; **Fox Chase Cancer Center** Douglas Flieder³¹, Jeff Boyd³¹, JoEllen Weaver³¹; **ILSbio** Bernard Kohl³², Dang Huy Quoc Thinh³²; **Indiana University** George Sandusky³³; **Indivumed** Hartmut Juhl³⁴; **John Flynn Hospital** Edwina Duhig^{35,36}; **Johns Hopkins University** Peter Illei⁶, Edward Gabrielson⁶, James Shin⁶, Beverly Lee⁶, Kristen Rogers⁶, Dante Trusty⁶, Malcolm V. Brock⁶; **Lahey Hospital & Medical Center** Christina Williamson³⁷, Eric Burks³⁷, Kimberly Rieger-Christ³⁷, Antonia Holway³⁷, Travis Sullivan³⁷; **Mayo Clinic** Dennis A. Wagle¹⁶, Michael K. Asiedu¹⁶, Farhad Kosari¹⁶; **Memorial Sloan-Kettering Cancer Center** William D. Travis⁴, Natasha Rehkman⁴, Maureen Zakowski⁴, Valerie W. Rusch⁴; **NYU Langone Medical Center** Paul Zippile³⁸, James Suh³⁸, Harvey Pass³⁸, Chandra Goparaju³⁸, Yvonne Owusu-Sarpong³⁸; **Ontario Tumour Bank** John M. S. Bartlett³⁹, Sugy Kodeeswaran³⁹, Jeremy Parfitt³⁹, Harmanjatinder Sekhon³⁹, Monique Albert³⁹; **Penrose St. Francis Health Services** John Eckman⁴⁰, Jerome B. Myers⁴⁰; **Roswell Park Cancer Institute** Richard Cheney⁴¹, Carl Morrison⁴¹, Carmelo Gaudio⁴¹; **Rush University Medical Center** Jeffrey A. Borgia⁴², Philip Bonomi⁴², Mark Pool⁴², Michael J. Liptay⁴²; **St. Petersburg Academic University** Fedor Moiseenko⁴³, Irina Zaytseva⁴³; **Thoraxklinik am Universitätsklinikum Heidelberg, Member of Biomaterial Bank Heidelberg (BMBH) & Biobank Platform of the German Centre for Lung Research (DZL)** Hendrik Dienemann⁴⁴, Michael Meister⁴⁴, Philipp A. Schnabel⁴⁵, Thomas R. Muley⁴⁴; **University of Cologne** Martin Peifer⁴⁶; **University of Miami** Carmen Gomez-Fernandez⁴⁷, Lynn Herbert⁴⁷, Sophie Egea⁴⁷; **University of North Carolina** Mei Huang¹¹, Leigh B. Thorne¹¹, Lori Boice¹¹, Ashley Hill Salazar¹¹, William K. Funkhouser¹¹, W. Kimryn Rathmell¹¹; **University of Pittsburgh** Rajiv Dhir⁴⁸, Samuel A. Yousem⁴⁸, Sanja Dacic⁴⁸, Frank Schneider⁴⁸, Jill M. Siegfried⁴⁸; **The University of Texas MD Anderson Cancer Center** Richard Hajek¹²; **Washington University School of Medicine** Mark A. Watson⁸, Sandra McDonald⁸, Bryan Meyers⁸; **Queensland Thoracic Research Center** Belinda Clarke³⁵, Ian A. Yang³⁵, Kwun M. Fong³⁵, Lindy Hunter³⁵, Morgan Windsor³⁵, Rayleen V. Bowman³⁵; **Center Hospitalier Universitaire Vaudois** Solange Peters⁴⁹, Igor Letovanec⁴⁹; **Ziauddin University Hospital** Khurram Z. Khan⁵⁰

Data Coordination Centre Mark A. Jensen⁵¹, Eric E. Snyder⁵¹, Deepak Srinivasan⁵¹, Ari B. Kahn⁵¹, Julien Baboud⁵¹, David A. Pot⁵¹

Project team: National Cancer Institute Kenna R. Mills Shaw⁵², Margi Sheth⁵², Tanja Davidsson⁵², John A. Demchok⁵², Liming Yang⁵², Zhining Wang⁵², Roy Tarnuzzer⁵², Jean Claude Zenklusen⁵²; **National Human Genome Research Institute** Bradley A. Ozenberger⁵³, Heidi J. Sofia⁵³

Expert pathology panel William D. Travis⁴, Richard Cheney⁴¹, Belinda Clarke³⁵, Sanja Dacic⁴⁸, Edwina Duhig^{36,35}, William K. Funkhouser¹¹, Peter Illei⁶, Carol Farver²⁷, Natasha Rehkman⁴, Gabriel Sica³⁰, James Suh³⁸ & Ming-Sound Tsao¹³

¹University of California San Francisco, San Francisco, California 94158, USA. ²The Eli and Edythe L. Broad Institute, Cambridge, Massachusetts 02142, USA. ³Dana Farber Cancer Institute, Boston, Massachusetts 02115, USA. ⁴Memorial Sloan-Kettering Cancer Center, New York, New York 10065, USA. ⁵University of Michigan, Ann Arbor, Michigan 48109, USA. ⁶Johns Hopkins University, Baltimore, Maryland 21287, USA. ⁷Baylor College of

Medicine, Houston, Texas 77030, USA. ⁸Washington University, St. Louis, Missouri 63108, USA. ⁹Harvard Medical School, Boston, Massachusetts 02115, USA. ¹⁰Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ¹¹University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA. ¹²University of Texas MD Anderson Cancer Center, Houston, Texas 77054, USA. ¹³Princess Margaret Cancer Centre, Toronto, Ontario M5G 2M9, Canada. ¹⁴Brigham and Women's Hospital Boston, Massachusetts 02115, USA. ¹⁵BC Cancer Agency, Vancouver, British Columbia V5Z 4S6, Canada. ¹⁶Mayo Clinic, Rochester, Minnesota 55905, USA. ¹⁷University of Southern California, Los Angeles, California 90033, USA. ¹⁸University of California Santa Cruz, Santa Cruz, California 95064, USA. ¹⁹Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. ²⁰University of Kentucky, Lexington, Kentucky 40515, USA. ²¹Buck Institute for Age Research, Novato, California 94945, USA. ²²Howard Hughes Medical Institute, University of California Santa Cruz, Santa Cruz, California 95064, USA. ²³Oregon Health and Science University, Portland, Oregon 97239, USA. ²⁴International Genomics Consortium, Phoenix, Arizona 85004, USA. ²⁵Analytical Biological Services, Inc., Wilmington, Delaware 19801, USA. ²⁶University of Alabama at Birmingham, Birmingham, Alabama 35294, USA. ²⁷Cleveland Clinic, Cleveland, Ohio 44195, USA. ²⁸Christiana Care, Newark, Delaware 19713, USA. ²⁹Cureline, Inc., South San Francisco, California 94080, USA. ³⁰Emory University, Atlanta, Georgia 30322, USA. ³¹Fox Chase Cancer Center, Philadelphia, Philadelphia 19111, USA. ³²ILSbio, Chestertown, Maryland 21620, USA. ³³Indiana University School of Medicine, Indianapolis, Indiana 46202, USA. ³⁴Indivumed, Silver Spring, Maryland 20910, USA. ³⁵The Prince Charles Hospital and the University of Queensland Thoracic Research Center, Brisbane, 4032, Australia. ³⁶Sullivan Nicolaides Pathology & John Flynn Hospital, Tugun 4680, Australia. ³⁷Lahey Hospital and Medical Center, Burlington, Massachusetts 01805, USA. ³⁸NYU Langone Medical Center, New York, New York 10016, USA. ³⁹Ontario Tumour Bank, Ontario Institute for Cancer Research, Toronto, Ontario M5G 0A3, Canada. ⁴⁰Penrose St. Francis Health Services, Colorado Springs, Colorado 80907, USA. ⁴¹Roswell Park Cancer Center, Buffalo, New York 14263, USA. ⁴²Rush University Medical Center, Chicago, Illinois 60612, USA. ⁴³St. Petersburg Academic University, St Petersburg 199034, Russia. ⁴⁴Thoraxklinik am Universitätsklinikum Heidelberg, 69126 Heidelberg, Germany. ⁴⁵University Heidelberg, 69120 Heidelberg, Germany. ⁴⁶University of Cologne, 50931 Cologne, Germany. ⁴⁷University of Miami, Sylvester Comprehensive Cancer Center, Miami, Florida 33136, USA. ⁴⁸University of Pittsburgh, Pittsburgh, Pennsylvania 15213, USA. ⁴⁹Center Hospitalier Universitaire Vaudois, Lausanne and European Thoracic Oncology Platform, CH-1011 Lausanne, Switzerland. ⁵⁰Ziauddin University Hospital, Karachi, 75300, Pakistan. ⁵¹SRA International, Inc., Fairfax, Virginia 22033, USA. ⁵²National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. ⁵³National Human Genome Research Institute, National Institutes of Health, Bethesda, Maryland 20892, USA.

Topoisomerase II mediates meiotic crossover interference

Liangran Zhang¹, Shunxin Wang¹, Shen Yin¹, Soogil Hong², Keun P. Kim² & Nancy Kleckner¹

Spatial patterning is a ubiquitous feature of biological systems. Meiotic crossovers provide an interesting example, defined by the classic phenomenon of crossover interference. Here we identify a molecular pathway for interference by analysing crossover patterns in budding yeast. Topoisomerase II plays a central role, thus identifying a new function for this critical molecule. SUMOylation (of topoisomerase II and axis component Red1) and ubiquitin-mediated removal of SUMOylated proteins are also required. The findings support the hypothesis that crossover interference involves accumulation, relief and redistribution of mechanical stress along the protein/DNA meshwork of meiotic chromosome axes, with topoisomerase II required to adjust spatial relationships among DNA segments.

During meiosis, crossovers promote genetic diversity and create physical connections between homologues that ensure their accurate segregation (reviewed in refs 1–3). Crossovers arise stochastically from a larger set of undifferentiated precursor recombination complexes, having different positions in different nuclei. Nonetheless, along any given chromosome in any given nucleus, crossovers tend to be evenly spaced (reviewed in refs 3 and 4). This genetic phenomenon is termed crossover interference^{5,6}.

Crossover interference implies the occurrence of communication along chromosomes, over distances ranging from 300 nm to more than 30 μm (refs 4, 7 and 8). Some models for crossover interference invoke spreading of a molecular-based change along the chromosomes⁹. Even spacing can also be achieved by a reaction–diffusion process¹⁰. We have proposed, alternatively, that interference involves the accumulation, relief and redistribution of mechanical stress, with spreading molecular changes following as a consequence of spreading stress relief⁴. Aberrant crossover patterns are observed in mutants defective for recombination, chromosome structure, chromatin state and DNA-based signal transduction. However, no specific molecular process has been defined. To address this deficit, we examined crossover patterns in wild-type (WT) and mutant strains of budding yeast as defined by cytological localization of crossover-correlated molecular foci.

Crossover interference in WT meiosis

Mammals, plants and fungi share a common meiotic recombination program. Recombination initiates by programmed double-strand breaks (DSBs), which occur as chromosome structural axes develop^{11,12}. Each DSB identifies its homologous partner duplex and mediates whole chromosome pairing. As a result, homologue structural axes are co-aligned, linked by bridging recombination complexes¹³. Crossover patterning is thought to act upon these bridging interactions^{13,14}, designating a subset to be crossovers, with accompanying interference^{14,15}. In yeast, crossover designation locally nucleates formation of synaptonemal complex between homologue axes^{13,14,16}. Synaptonemal complex then spreads along the lengths of the chromosomes. Correspondingly, crossover patterning and interference are independent of synaptonemal complex formation^{13,17,18} (below).

In yeast, foci of E3 ligase Zip3, which specifically mark the sites of patterned crossovers^{8,18–20}, serve as an early marker for crossover interference analysis (Methods). Zip3 foci emerge immediately following crossover designation, thus avoiding complications arising during

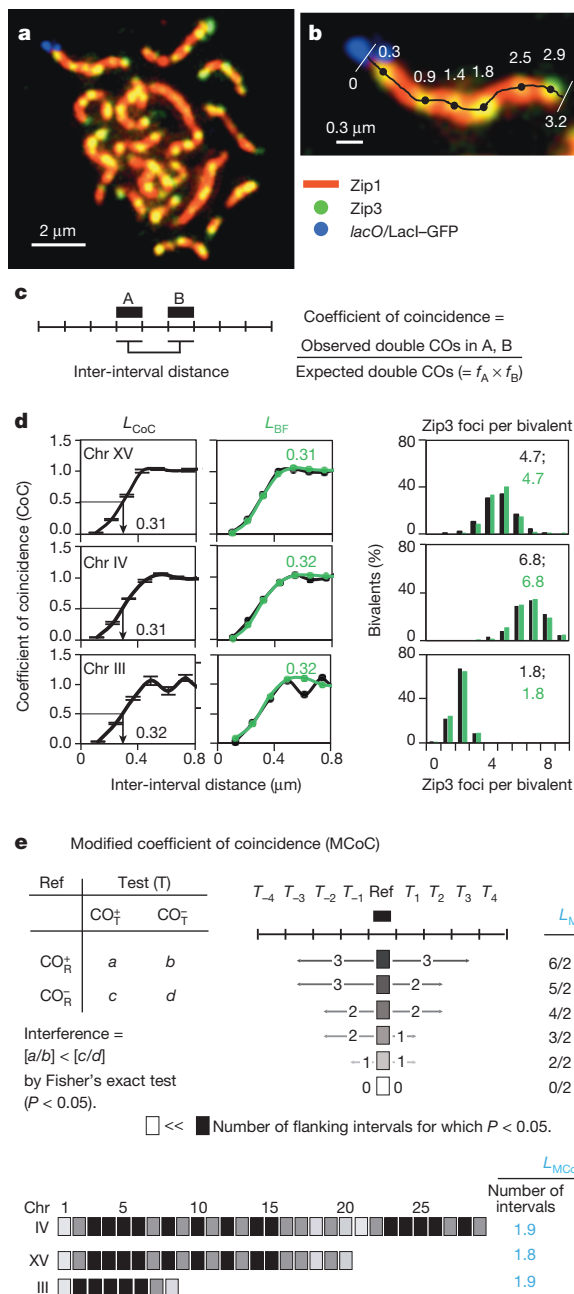
formation of actual crossover products⁸. Also, Zip3 foci do not mark the sites of additional crossovers that arise by other routes⁸ (Methods).

Zip3-MYC foci were visualized along synaptonemal complexes of surface-spread pachytene chromosomes by wide-field epifluorescence⁸ (Fig. 1a, b). Each Zip3 focus position was defined, to an accuracy of approximately one pixel (67 nm) along a particular marked chromosome in ~ 200 –300 nuclei, thus defining patterns with a high degree of reproducibility and accuracy⁸ (Supplementary Table 1). Using these position data, the distance along a chromosome over which the interference signal is detectable (that is, the ‘interference distance’, L) is defined by three different approaches (Fig. 1c–e). In each case, L is given in units of physical distance (rationale below), micrometres of synaptonemal complex, which is a proxy for chromosome length at late leptotene when crossover designation actually occurs (above).

Crossover interference is classically described by coefficient of coincidence (CoC) analysis^{5,6,8} (Fig. 1c; see Methods). Zip3 foci along three chromosomes of different sizes (330–1,530 kilobases (kb)) exhibit classic coefficient of coincidence relationships (Fig. 1d, left column). For intervals that are close together, bivalents exhibiting a focus in each interval (‘double events’) are much rarer than expected, reflecting operation of interference; as the inter-interval distance increases, double-event frequencies progressively approach, then reach, that expected for independent occurrence, where the observed frequency is the same as the expected frequency (coefficient of coincidence = 1). At even longer intervals, coefficient of coincidence values can exceed 1, reflecting the tendency for even spacing⁸. For convenience, we define the interference distance described by such curves as the inter-interval distance at which the coefficient of coincidence = 0.5, namely L_{CoC} (Fig. 1d, left column). The three analysed chromosomes exhibit virtually identical coefficient of coincidence curves and values of $L_{\text{CoC}} = 0.3 \pm 0.01 \mu\text{m}$ ($n = 2$ –4; Fig. 1d, left column; Methods).

We previously described a stress-and-stress relief mechanism for crossover patterning (the ‘beam-film’ (BF) model). Beam-film-predicted crossover patterns are defined by simulation analyses⁸ that can accurately describe crossover patterns in diverse organisms, including yeast (Fig. 1d, middle and right columns). The beam-film parameter (L) is the distance over which the interference signal spreads along the chromosomes and corresponds to the distance at which the predicted coefficient of coincidence = 0.5, namely L_{BF} . Beam-film simulations give the same value of L and $L_{\text{BF}} \approx 0.3 \mu\text{m}$ for all three analysed yeast chromosomes (Fig. 1d, middle column).

¹Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ²Department of Life Science, Chung-Ang University, Seoul 156-756, South Korea.



Crossover interference can be examined by a modified coefficient of coincidence analysis (MCoC²¹, Fig. 1e; Methods). The three analysed yeast chromosomes exhibit the same average L_{MCoC} of $\sim 0.3 \mu\text{m}$.

Crossover interference requires topoisomerase II

Topoisomerase II (TopoII) alleviates topological stresses within chromosomes. If crossover interference involves mechanical stress along the chromosomes⁴, TopoII could be a key player. We assessed crossover interference in three mutants with altered TopoII states (Fig. 2 and Extended Data Figs 1–3). (1) TopoII was depleted using a *pCLB2-TOP2* fusion that expressed TopoII in vegetative cells but not meiosis. (2) TopoII catalytic activity was eliminated in meiosis by expressing a catalytically inactive allele (*top2YF*) under its native promoter in a *pCLB2-TOP2* strain, leaving *top2YF* as the only gene expressed during meiosis. (3) SUMOylation of TopoII at several carboxy (C)-terminal residues²² was eliminated by mutation. All three *top2* mutant strains grow well vegetatively, progress to the pachytene stage of meiosis and

exhibit normal synaptonemal complex morphology and length²³ (Extended Data Fig. 3). Meiotic TopoII levels and localization were severely reduced in *pCLB2-TOP2* and not detectably changed in other mutants (Extended Data Fig. 1). In all three *top2* mutant strains, for all three analysed chromosomes, the interference distance as defined by L_{CoC} , L_{BF} and L_{MCoC} decreased from $\sim 0.3 \mu\text{m}$ in WT to $\sim 0.2 \mu\text{m}$ (Fig. 2a, b and Extended Data Figs 2 and 3). Reduced interference should be accompanied by an increased number of crossovers, and, in all cases, the distribution of Zip3 foci per bivalent was shifted to higher values (Fig. 2a, b and Extended Data Fig. 2).

For *pCLB2-TOP2*, the interference defect was confirmed by a fourth approach. Meiotic crossover patterns are characterized by 'crossover homeostasis'²⁴. A decrease or increase in the frequency of DSBs (and thus crossover precursor interactions) necessarily changes the crossover frequency. However, the magnitudes of such changes are less than proportional to the change in DSB/precursor frequency, implying a homeostatic effect. Crossover homeostasis is a direct consequence of crossover interference^{8,24}: homeostatic disparity is greater or less when crossover interference is stronger or weaker, and absent when crossover interference is absent. This interplay is predicted, and can be quantified, by beam-film simulations⁸ (Fig. 2d).

To evaluate crossover homeostasis experimentally, the number of Zip3 foci along a given chromosome was determined in a series of strains that exhibited different levels of DSBs (precursors). Decreased and increased levels are conferred by hypomorphic mutations in DSB transesterase Spo11 and a *tel1A* mutation respectively⁸ (Extended Data Fig. 4 and Fig. 2d). In a *TOP2* background, homeostasis is apparent in the nonlinear relationship of Zip3 focus number to DSB number (chromosomes XV and III; Fig. 2d, filled black circles, and Extended Data Fig. 4)⁸. Moreover, the experimentally defined relationships occur at exactly the level of interference predicted to occur in WT meiosis by best-fit beam-film simulation analysis⁸ ($L_{BF} \approx 0.3 \mu\text{m}$; above; Fig. 2d).

If *pCLB2-TOP2* reduces the interference distance, it should bring the relationship between Zip3 focus number and DSB number closer to the linear proportionality seen in the absence of interference. This prediction is fulfilled (chromosomes XV and III; Fig. 2d filled pink circles and Extended Data Fig. 4). Furthermore, the mutant relationships again occur specifically at the interference distance predicted by best-fit beam-film simulation analysis for this mutant ($L_{BF} \approx 0.2 \mu\text{m}$; Fig. 2d and Extended Data Fig. 4). These results confirm the existence of an interference defect in *pCLB2-TOP2* and provide further evidence that the beam-film model can accurately describe crossover patterns (see also Extended Data Fig. 4).

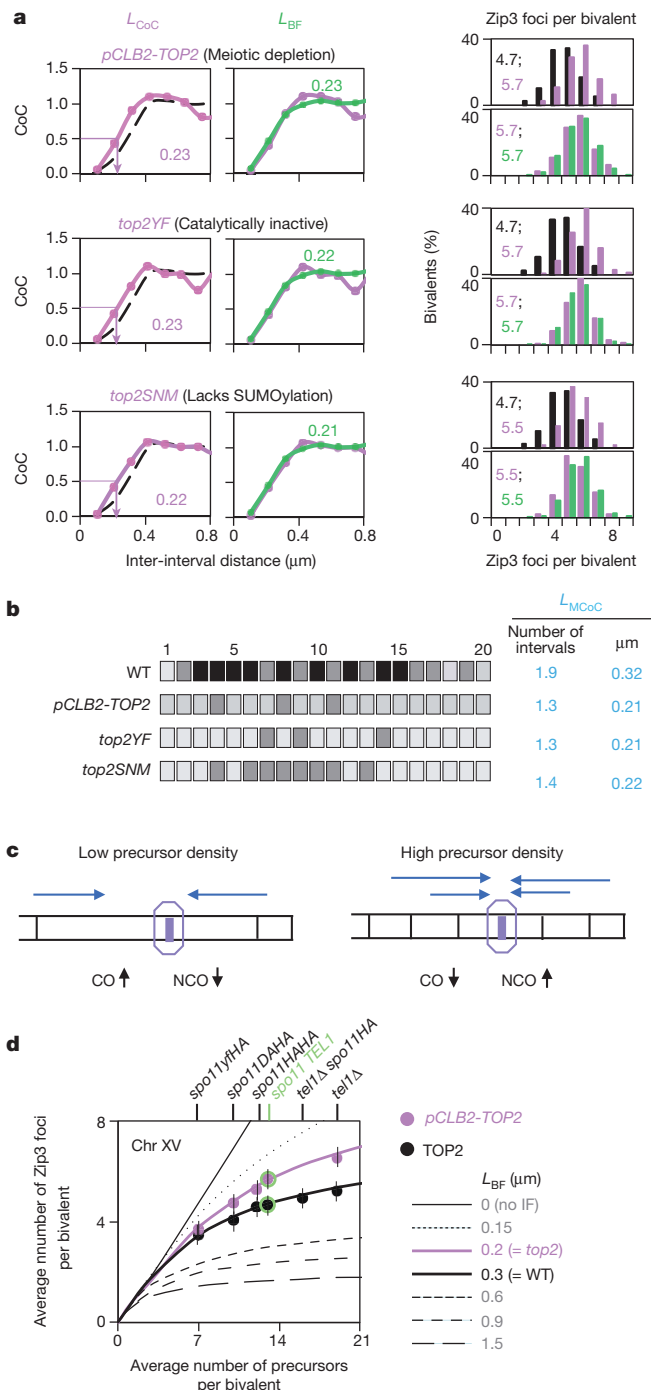


Figure 2 | Crossover interference in *top2* mutants. **a, b**, All three *top2* mutants show decreased crossover interference by all criteria (L_{CoC} , L_{BF} , L_{MCoC}) and correspondingly increased crossover frequency. **a**, WT, *top2* and beam-film (BF) simulation data (black, pink and green). **c**, The basis for crossover homeostasis⁸. CO, crossover; NCO, non-crossover. At lower (higher) precursor density (black vertical lines; left (right)), a given precursor will be less (more) likely to experience interference emanating from nearby crossovers (indicated by fewer (more) blue lines), giving an increased (decreased) probability of a crossover at each individual position, and thus along the whole chromosome length. The magnitudes of these effects will be greater or lesser according to the strength of crossover interference (and zero in its absence). **d**, Quantitative evaluation of crossover homeostasis on chromosome XV. Lines: relationship of crossover number to precursor number (parameter N) predicted by beam-film simulations at varying interference levels (L_{BF} = interference distance, L ; other parameters appropriate to WT yeast meiosis⁸). Crossover homeostasis decreases with decreasing crossover interference. Filled circles: strains exhibiting altered DSB levels (top) were analysed for Zip3 foci in *TOP2* (black) and *pCLB2-TOP2* (pink) backgrounds (Extended Data Fig. 4). Average frequency of Zip3 foci per bivalent plotted versus DSB (= precursor) number (vertical lines indicate s.d.). *pCLB2-TOP2* differs experimentally from WT in the direction expected for decreased crossover interference. Experimental data for WT and *pCLB2-TOP2* both quantitatively match the relationships predicted for their corresponding interference levels by beam-film simulations (L_{BF} = 0.3 and 0.2 μm , respectively; Fig. 2a, b).

of either the Slx5 SUMO-binding motif or the Slx8 ubiquitin ligase motif (*slx5A*, *slx8A*, *slx5-SIM* or *slx8-SS*), confers the same changes in Zip3 focus patterns as *top2*, *red1-KR* and *ubc9-GFP* (Fig. 3b and Extended Data Figs 2 and 3). The *slx5A* defect is confirmed genetically (Extended Data Fig. 3 and Supplementary Table 2).

The sirtuin, Sir2, enables Slx5/8 STUbL activity²⁹ and is required for crossover interference via that activity. The absence of Sir2 (*sir2A*) or specific elimination of the interaction of Sir2 with Slx5/8 (*sir2RK*) confer the same changes in Zip3 focus patterns as all of the other mutations analysed above, by all criteria (Fig. 3c and Extended Data Figs 2 and 3). The interference defect in *sir2RK* was confirmed genetically (Extended Data Fig. 3 and Supplementary Table 2).

The role of Sir2 in interference is specific to this one function. Elimination of other Sir2-mediated activities (histone deacetylase catalysis (*sir2-345*); interaction with Sir2 partners required for silencing (deletion mutants of Sir3, Sir4, Esc2 and Esc8); cohesion (*sir2AC500*)) does not alter crossover interference (Fig. 3c and Extended Data Figs 3 and 6).

A single TopoII crossover interference pathway

Not only do all analysed mutants exhibit the same quantitative defects in crossover interference and crossover number as defined by Zip3 focus patterns (Figs 2–4 and Extended Data Figs 2 and 3), but double mutants carrying combinations of single mutations also exhibit these same phenotypes (Fig. 4a,b). Thus, the described mutants define a single molecular pathway.

This pathway may directly implement the spreading interference signal, but other perturbations are not excluded (Supplementary Discussion). These results cannot be explained by (1) prolongation of the crossover-designation period, (2) higher DSB/precursor levels (Extended Data Figs 4, 7 and 8) or (3) obviously altered axis organization, since all mutants exhibit WT synaptonemal complex lengths (Extended Data Fig. 3). All mutants exhibit reduced evenness of spacing as defined by gamma distribution analysis (Supplementary Discussion).

The obligatory crossover does not require interference

Since a crossover is required for meiotic homologue segregation, every pair of homologues must acquire at least one (the ‘obligatory crossover’)³. The frequency of zero-Zip3 focus chromosomes is less than 10^{-3} for chromosomes IV and XV and $\sim 1\%$ for chromosome III because it is small⁸. None of the identified interference-defective mutants exhibits an increased frequency of zero-Zip3 foci chromosomes (Figs 1–4 and

Crossover interference requires SUMO and STUbL

SUMOylation of TopoII requires Ubc9, the only known SUMO–E2 of yeast²⁵. Another Ubc9 substrate is meiotic axis component Red1 (ref. 20). Mutation of the SUMOylation patch of Red1, which dramatically reduces the level of modification (*red1KR*²⁶), confers the same altered Zip3 focus patterns as *top2* mutations, including *top2SNM* (Fig. 3a and Extended Data Fig. 3). Interestingly the non-null allele, *ubc9-GFP*²⁷, also exhibits this phenotype (Fig. 3a and Extended Data Fig. 3), as well as an elevated level of crossovers as defined genetically²⁷.

Crossover interference also requires STUbL protein Slx5/8. Slx5/8 ubiquitinates SUMOylated proteins, targeting them for removal from their cognate complexes²⁸. Absence of Slx5/8 activity confers a strong global increase in protein SUMOylation during meiosis (Extended Data Fig. 5). Absence of either Slx5 or Slx8, or mutational abrogation

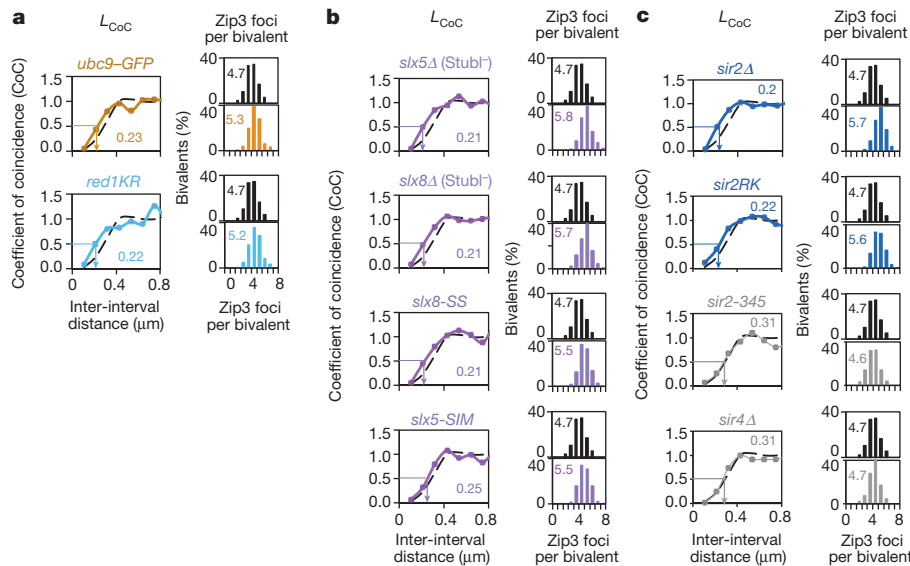


Figure 3 | Crossover interference requires post-translational modification. **a–c**, WT and mutant crossover patterns (black; colours). Quantitatively similar decreases in crossover interference and increases in crossover number are seen in *ubc9-GFP* (SUMO E2; brown), *red1KR* (non-SUMOylated Red1; cyan), strains lacking Slx5 or Slx8 (*slx5Δ* or *slx8Δ*) or mutated for the Slx5 SUMO-binding

motif or the Slx8 ubiquitin ligase motif (*slx5-SIM*, *slx8-SS*) (magenta), or lacking Sir2 (*sir2Δ*) or mutated for the Sir2/Slx5 interaction site (*sir2RK*) (blue). Crossover interference does not require Sir2 deacetylation activity (*sir2-345*), Sir2 interaction partner Sir4 (*sir4Δ*) (grey) or other Sir2 activities/partners (text).

Extended Data Fig. 2). This result argues against models in which crossover interference is required to ensure the obligatory crossover^{8,9} whereas the beam-film model predicts this phenotype⁸.

The crossover interference metric is physical distance

We analysed Zip3 focus patterns in strains whose pachytene synaptonemal complex lengths differ from those of the reference WT SK1 strain (Fig. 5 and Extended Data Fig. 9). These strains exhibit different interference distances when the metric used is genomic length (kilobases) but exactly the same (WT) interference distance when the metric is physical length as micrometres of synaptonemal complex (Fig. 5; compare top and bottom panels). Beam-film simulations give the same relationships (Extended Data Fig. 9a–c). Thus, in budding yeast, the metric for spreading crossover interference is physical chromosome distance, as in mouse, *Arabidopsis*, human and tomato^{8,30–32}. Differences in synaptonemal complex length probably result from altered chromatin loop lengths (kilobases) without a change in basic axis structure^{33,34}. In all cases, experimental Zip3 focus distributions are matched by beam-film simulations that use the WT value for interference distance (L_{BF}). These and other details (Extended Data Fig. 9 legend) provide further evidence of the precision with which the beam-film model explains diverse crossover patterns.

The TopoII interference pathway is highly specific

None of more than 20 other examined mutants exhibit altered Zip3 focus patterns, including those with the following: (1) altered axis composition (condensin, *pch2Δ*); (2) lacking either a sister chromatid (*cdc6*) or any/normal synaptonemal complex (*zip1Δ*; *msh4Δ*)^{8,18} (Fig. 5; discussion in Extended Data Fig. 9a and Methods); or (3) deleted for Sir2 relative, Hst1; ATM homologue Tel1; meiotic telomere/motion protein Ndj1; chromodomain protein Dot1; DSB-triggered γ -H2A; TopoII-co-localizing Nse1/Smc5/6; nucleosome density factor Yta7; Mph1, Mlh1/3 and Mms4 (recombination resolution); or Msh2 (mismatch repair) (Extended Data Fig. 6; L.Z., unpublished observations).

Discussion

Our findings show that Topoisomerase II is essential for normal CO interference. Further, crossover interference is mediated by communication

along prophase chromosome structural axes (Fig. 6a). The TopoII interference pathway involves SUMOylation of Red1, a prominent meiotic axis component. TopoII itself occurs prominently along meiotic prophase axes, in yeast and mammals^{35,36} and along the structural axes of mammalian mitotic late-stage chromosomes, to which meiotic axes are related³⁷. Moreover, the TopoII interference pathway requires SUMOylation of TopoII as well as Red1. In mitotic mammalian cells, SUMOylated TopoII is implicated in late-stage chromosome structural axes³⁸, and in yeast, SUMOylated TopoII occurs preferentially in centromere regions³⁹ which, during meiosis, mimic crossover-designation/interference sites by nucleating synaptonemal complex formation¹⁶. Spreading of interference along the axis matches our finding that the relevant metric is physical chromosome distance and the inference that variations in synaptonemal complex length in different mutants result from variations in loop length rather than basic axis structure. Finally, spreading along the axis explains how the interference signal is first generated by, then sensed by, biochemical recombination complexes, which are intimately embedded in the axes from their first inception as pre-DSB ensembles¹². Notably, the meiotic prophase axis probably comprises a meshwork of DNA segments joined by linker proteins^{1,33,37} (Fig. 6a, b).

Most importantly, crossover interference requires the catalytic activity of TopoII. Since TopoII activity does not require input of external energy from ATP hydrolysis, its reactions must be driven forward, and given directionality, by their substrates, which are changed by TopoII from a higher potential energy state to a lower potential energy state. If substrate for TopoII during crossover interference is the axis meshwork (above), that meshwork is first placed in a high potential energy state and then, in response to crossover designation, undergoes relaxation, dependent upon TopoII activity. That is, the axis meshwork begins in a mechanically stressed state and is then relaxed to a less mechanically stressed state dependent upon TopoII. This progression closely matches the proposed stress and stress relief mechanism for crossover patterning^{4,14} (Methods): stress accumulates along the chromosomes and provokes local crossover designation which, by its intrinsic nature, results in local relief of stress. That local change then redistributes along the chromosomes, emanating outwards from its nucleation site, reducing stress and thereby disfavours additional stress-promoted crossover designations in the affected regions.

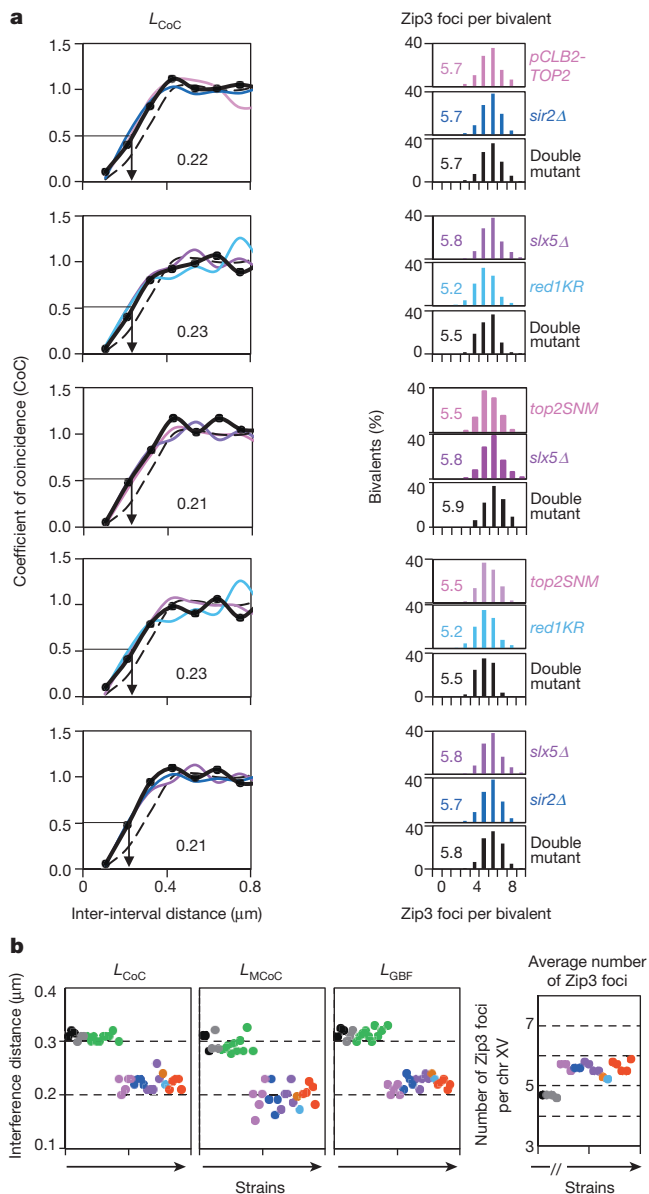


Figure 4 | A single pathway for crossover interference. **a**, Representative double mutants and component single mutants exhibit the same quantitative defect in crossover interference and increased crossover number (colours and black) versus WT (dashed line). **b**, Crossover interference and crossover number phenotypes for all mutants (Figs 1–5). WT (black), *sir2-345* and *sir4 Δ* (grey), *top2* mutants (pink), *sir2 Δ* and *sir2RK* (blue), *slx5/8* mutants (purple), *ubc9-GFP* (brown), *red1KR* (light blue), double mutants (red); mutants with altered axis length showing WT phenotype (green).

In this context, what is the source of meshwork stress and how does TopoII alleviate that stress? We previously suggested that mechanical stress arises from axis-constrained global chromatin expansion; cross-over designation and interference then involve local nucleation and spreading of chromatin/axis compaction^{4,14} (Fig. 6b). TopoII could act during compaction to adjust spatial relationships among DNA segments within the axis meshwork (Fig. 6b), thereby implementing both local relief of stress and its redistribution. The stress-relief role of TopoII is thus specifically targeted to the compaction process, and thus to regions undergoing crossover designation/interference. This role also explains why the TopoII pathway is important, but not absolutely essential, for crossover interference: in its absence, the basic process of spreading stress relief would occur, but full relaxation would not be possible without meshwork readjustment (Fig. 6b). Interestingly, mitotic chromosomes are constrained by topologically sensitive linkages and collapse upon removal of protein/DNA links^{40,41}, exactly as expected for a meshwork under expansion stress.

We further note that the beam-film model, formulated to quantitatively describe the predictions of a stress and stress-relief mechanism^{4,8}, accurately and quantitatively describes diverse crossover patterning data for WT meiosis, including crossover homeostasis, in yeast and other organisms⁸ (Figs 1d and 2d), as well as crossover patterning in mutants. These include the following: (1) crossover interference, crossover number and crossover homeostasis in mutants defective in the TopoII interference pathway (Fig. 2a, d; not shown); (2) crossover patterns at varying DSB levels in those mutants (Extended Data Fig. 4); and (3) crossover patterns in mutants with altered axis lengths (Extended Data Fig. 9a, b). Recent findings in *Caenorhabditis elegans*⁴² can also be directly explained by such a model (Supplementary Discussion). Importantly, however, the mathematical formulation of the beam-film model can equivalently describe any mechanism involving progressive ‘event designation’ and resulting interference that decays exponentially away from the designation site. Thus, proof that crossover patterning involves macroscopic mechanical effects requires direct identification of such effects.

Finally, our results implicate SUMOylation (of Red1 and TopoII, probably among multiple targets) and ubiquitin-targeted removal of SUMOylated proteins in the TopoII crossover interference pathway. These effects presumably act sequentially on the same molecules, which are first specifically SUMOylated and then targeted for removal via STUbL activity. SUMOylation might establish preconditions for cross-over interference whose subsequent implementation would require removal of those SUMOylated proteins. Alternatively, SUMOylation and STUbL activity might compete actively in a single aspect of the patterning process; or SUMOylation might function only to target protein removal. For yeast TopoII, absence of SUMOylation (in *top2SNM*) decreases the mobility of chromosome-bound TopoII⁴³, perhaps promoting repeated cycles of TopoII catalytic activity.

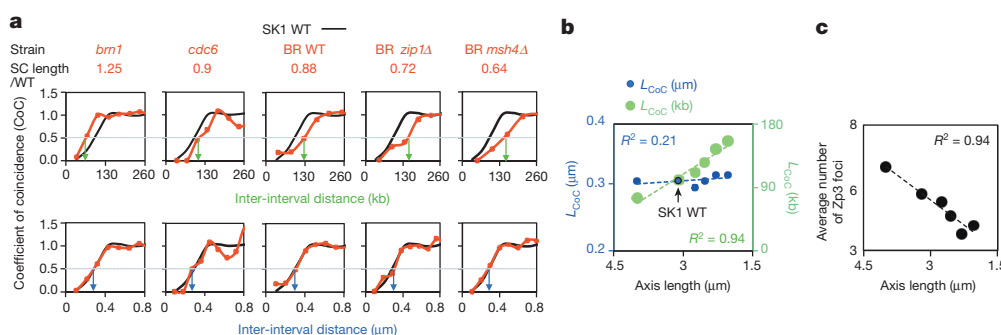


Figure 5 | The metric of crossover interference is physical chromosomal length (micrometres). **a**, **b**, Coefficient of coincidence relationships for strains with different axis lengths (red) relative to WT SK1 (black). **a**, Interference lengths differ in genomic distance (kilobases) (top) but are the same in physical distance (micrometres of synaptonemal complex (SC)) (bottom). **b**, L_{CoC} values from **a** with corresponding linear regression lines. **c**, Zip3 focus frequencies vary linearly with bivalent axis/synaptonemal complex length.

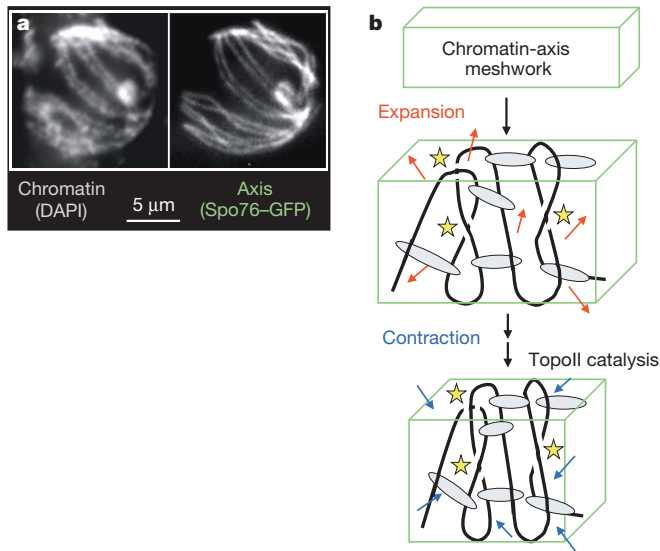


Figure 6 | Proposed role of TopoII for crossover interference.

a, Chromosomes at the crossover-designation stage (late leptotene), visualized in *Sordaria*, suggest that the axis (identified with Spo76-GFP) incorporates a significant fraction of chromatin (stained with DAPI) in a DNA/protein structural meshwork. **b**, Model. Top: global chromatin expansion within the structural axis meshwork is constrained by meshwork tethers, giving an expanded, mechanically stressed meshwork state. Bottom: spreading interference creates a more contracted state with resulting reduction in mechanical meshwork stress. Full implementation of contraction, and thus maximal spreading of interference, requires readjustment of spatial relationships among component DNA segments which, comprising topologically closed domains, require TopoII-mediated duplex/duplex passages (yellow stars).

METHODS SUMMARY

Analysed yeast strains were isogenic SK1 derivatives (Extended Data Table 1). Details of analyses are described in ref. 8, Methods and Extended Data. Zip3 focus positions and synaptonemal complex lengths for all experiments are in Supplementary Table 1. Data for BR strains (Fig. 5) were provided by J. Fung.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 November 2013; accepted 2 May 2014.

Published online 13 July 2014.

- Kleckner, N., Zhang, L., Weiner, B. & Zickler, D. in *Genome Organization and Function in the Cell Nucleus* (ed. Rippe, K.) 487–533 (John Wiley, 2011).
- Zickler, D. & Kleckner, N. The leptotene-zygotene transition of meiosis. *Annu. Rev. Genet.* **32**, 619–697 (1998).
- Jones, G. H. & Franklin, F. C. Meiotic crossing-over: obligation and interference. *Cell* **126**, 246–248 (2006).
- Kleckner, N. *et al.* A mechanical basis for chromosome function. *Proc. Natl Acad. Sci. USA* **101**, 12592–12597 (2004).
- Muller, H. J. The mechanism of crossing over, parts I–IV. *Am. Nat.* **50**, 193–434 (1916).
- Sturtevant, A. H. The behavior of the chromosomes as studied through linkage. *Z. indukt. Abstamm.-u. VererbLehre* **13**, 234–287 (1915).
- Hillers, K. J. & Villeneuve, A. M. Chromosome-wide control of meiotic crossing over in *C. elegans*. *Curr. Biol.* **13**, 1641–1647 (2003).
- Zhang, L., Liang, Z., Hutchinson, J. & Kleckner, N. Crossover patterning by the Beam-Film model: analysis and implications. *PLoS Genet.* **10**, e1004042 (2014).
- King, J. S. & Mortimer, R. K. A polymerization model of chiasma interference and corresponding computer simulation. *Genetics* **126**, 1127–1138 (1990).
- Vecchiarelli, A. G., Hwang, L. C. & Mizuuchi, K. Cell-free study of F plasmid partition provides evidence for cargo transport by a diffusion-ratchet mechanism. *Proc. Natl Acad. Sci. USA* **110**, E1390–E1397 (2013).
- Blat, Y., Protacio, R. U., Hunter, N. & Kleckner, N. Physical and functional interactions among basic chromosome organizational features govern early steps of meiotic chiasma formation. *Cell* **111**, 791–802 (2002).
- Pan, J. *et al.* A hierarchical combination of factors shapes the genome-wide topography of yeast meiotic recombination initiation. *Cell* **144**, 719–731 (2011).
- Storlazzi, A. *et al.* Recombination proteins mediate meiotic spatial chromosome organization and pairing. *Cell* **141**, 94–106 (2010).

- Borner, G. V., Kleckner, N. & Hunter, N. Crossover/noncrossover differentiation, synaptonemal complex formation, and regulatory surveillance at the leptotene/zygotene transition of meiosis. *Cell* **117**, 29–45 (2004).
- Hunter, N. in *Molecular Genetics of Recombination, Topics in Current Genetics* (eds Aguilera, A. & Rothstein, R.) 381–442 (Springer, 2006).
- Henderson, K. A. & Keeney, S. Synaptonemal complex formation: where does it start? *Bioessays* **27**, 995–998 (2005).
- Bishop, D. K. & Zickler, D. Early decision; meiotic crossover interference prior to stable strand exchange and synapsis. *Cell* **117**, 9–15 (2004).
- Fung, J. C., Rockmill, B., Odell, M. & Roeder, G. S. Imposition of crossover interference through the nonrandom distribution of synapsis initiation complexes. *Cell* **116**, 795–802 (2004).
- Agarwal, S. & Roeder, G. S. Zip3 provides a link between recombination enzymes and synaptonemal complex proteins. *Cell* **102**, 245–255 (2000).
- Cheng, C. H. *et al.* SUMO modifications control assembly of synaptonemal complex and polycomplex in meiosis of *Saccharomyces cerevisiae*. *Genes Dev.* **20**, 2067–2081 (2006).
- Malkova, A. *et al.* Gene conversion and crossing over along the 405-kb left arm of *Saccharomyces cerevisiae* chromosome VII. *Genetics* **168**, 49–63 (2004).
- Bachant, J., Alcasabas, A., Blat, Y., Kleckner, N. & Elledge, S. J. The SUMO-1 isopeptidase Smt4 is linked to centromeric cohesion through SUMO-1 modification of DNA topoisomerase II. *Mol. Cell* **9**, 1169–1182 (2002).
- Rose, D. & Holm, C. Meiosis-specific arrest revealed in DNA topoisomerase II mutants. *Mol. Cell. Biol.* **13**, 3445–3455 (1993).
- Martini, E., Diaz, R. L., Hunter, N. & Keeney, S. Crossover homeostasis in yeast meiosis. *Cell* **126**, 285–295 (2006).
- Baldwin, M. & Bachant, J. Top2 SUMO conjugation in yeast cell lysates. *Methods Mol. Biol.* **582**, 209–219 (2009).
- Eichinger, C. S. & Jentsch, S. Synaptonemal complex formation and meiotic checkpoint signaling are linked to the lateral element protein Red1. *Proc. Natl Acad. Sci. USA* **107**, 11370–11375 (2010).
- Hooker, G. W. & Roeder, G. S. A Role for SUMO in meiotic chromosome synapsis. *Curr. Biol.* **16**, 1238–1243 (2006).
- Nagai, S., Davoodi, N. & Gasser, S. M. Nuclear organization in genome stability: SUMO connections. *Cell Res.* **21**, 474–485 (2011).
- Darst, R. P., Garcia, S. N., Koch, M. R. & Pillus, L. Slx5 promotes transcriptional silencing and is required for robust growth in the absence of Sir2. *Mol. Cell. Biol.* **28**, 1361–1372 (2008).
- Drouaud, J. *et al.* Sex-specific crossover distributions and variations in interference level along *Arabidopsis thaliana* chromosome 4. *PLoS Genet.* **3**, e106 (2007).
- Petkov, P. M., Broman, K. W., Szatkiewicz, J. P. & Paigen, K. Crossover interference underlies sex differences in recombination rates. *Trends Genet.* **23**, 539–542 (2007).
- Hou, Y. *et al.* Genome analyses of single human oocytes. *Cell* **155**, 1492–1506 (2013).
- Kleckner, N. Chiasma formation: chromatin/axis interplay and the role(s) of the synaptonemal complex. *Chromosoma* **115**, 175–194 (2006).
- Novak, I. *et al.* Cohesin Smc1beta determines meiotic chromatin axis loop organization. *J. Cell Biol.* **180**, 83–90 (2008).
- Klein, F. *et al.* Localization of RAP1 and topoisomerase II in nuclei and meiotic chromosomes of yeast. *J. Cell Biol.* **117**, 935–948 (1992).
- Moens, P. B. & Earnshaw, W. C. Anti-topoisomerase II recognizes meiotic chromosome cores. *Chromosoma* **98**, 317–322 (1989).
- Kleckner, N., Zickler, D. & Witz, G. Chromosome capture brings it all together. *Science* **342**, 940–941 (2013).
- Agostinho, M. *et al.* Conjugation of human topoisomerase 2α with small ubiquitin-like modifiers 2/3 in response to topoisomerase inhibitors: cell cycle stage and chromosome domain specificity. *Cancer Res.* **68**, 2409–2418 (2008).
- Lee, M. T. & Bachant, J. SUMO modification of DNA topoisomerase II: trying to get a CENse of it all. *DNA Repair* **8**, 557–568 (2009).
- Kawamura, R. *et al.* Mitotic chromosomes are constrained by topoisomerase II-sensitive DNA entanglements. *J. Cell Biol.* **188**, 653–663 (2010).
- Pope, L. H., Xiong, C. & Marko, J. F. Proteolysis of mitotic chromosomes induces gradual and anisotropic decondensation correlated with a reduction of elastic modulus and structural sensitivity to rarely cutting restriction enzymes. *Mol. Biol. Cell* **17**, 104–113 (2006).
- Libuda, D. E., Uzawa, S., Meyer, B. J. & Villeneuve, A. M. Meiotic chromosome structures constrain and respond to designation of crossover sites. *Nature* **502**, 703–706 (2013).
- Warsi, T. H. in *Centromeric Functions and Dynamics of DNA Topoisomerase II in S. cerevisiae* 130–187. Ph.D. thesis, Univ. California Riverside (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Hochstrasser, J. Bachant, S. Jentsch, L. Pillus and M. Weinreich for plasmids, J. Fung for Zip2 focus data, D. Zickler for the image in Fig. 6a, and members of the Kleckner laboratory and D. Zickler for advice and discussions. This research, L.Z., S.W., S.Y. and N.K. were supported by a grant to N.K. from the National Institutes of Health (R01 GM044794); S.H. and K.P.K. were supported by the National Research Foundation of Korea funded by the Ministry of Science, ICT and Future Planning (2012-M3A9C6050367).

Author Contributions L.Z. and N.K. conceived and designed experiments, analysed data and wrote the paper. L.Z., S.W., Y.S., S.H. and K.P.K. performed experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.K. (kleckner@fas.harvard.edu).

METHODS

Strains. Yeasts strains are isogenic derivatives of SK1 (Extended Data Table 1) except for BR strains (Fig. 5), for which Zip2 foci data were provided by J. Fung (ref. 18).

Pachytene Zip2/Zip3 foci mark the sites of patterned ('interfering') crossovers. In budding yeast, as in many organisms, the majority of crossovers arise as the consequence of the programmed patterning process characterized by crossover interference. However, a minority of crossovers arise in some other way. The two types of crossover are referred to as 'patterned', 'class I' or 'interfering', and as 'class II' or 'non-interfering', respectively. We prefer to avoid the terms 'interfering' and 'non-interfering' for reasons discussed below.

There are a total of approximately 90 crossovers per yeast nucleus per round of meiosis as defined both by microarray and genetic analyses^{44–46}. Mutant analysis suggests that the patterned (class I) crossovers constitute about 70% of total crossovers (estimates range from 60% to 90% in different studies; see, for example, refs 47, 48). Approximately 70% of about 90 total crossovers implies about 63 patterned (class I) crossovers per nucleus.

Zip2/3 foci appear to specifically mark the sites of patterned (class I) crossovers by several criteria, as follows.

First, there are approximately 65 foci of Zip2, Zip3 and Msh4/5 on yeast pachytene chromosomes per nucleus, and these different types of focus are highly co-localized with one another, implying that they mark the same specific set of recombinational interactions^{8,19,49–51}. These foci also co-localize with DSBs formation/repair components, for example Mre11 and Rad51/Dmc1, implying that they mark the sites of recombinational interactions (see, for example, refs 18, 19, 50; L.Z., unpublished observations). The number of these foci corresponds well with the predicted number of patterned crossovers (above). Furthermore, crossover levels defined genetically co-vary with the number of Zip2/3 and Msh4/5 foci in mutants examined, for example *sgs1Δ*, *tel1Δ* and *spo11* hypomorphs, implying that they represent an important majority of recombinational interactions (refs 24, 52, 53 and this study). Additionally, Zip2/3 and Msh4/5 have all been implicated specifically in maturation of patterned/interfering crossovers (see, for example, refs 8, 18, 19, 44, 50, 51).

Second, Zip2 and Zip3 foci exhibit robust interference as shown both by coefficient of coincidence relationships for random adjacent pairs of intervals and by full coefficient of coincidence relationships along specific individual chromosomes (refs 8, 18 and this study). Also, the number of Zip3 foci shows crossover homeostasis as defined in strains with altered DSB levels (refs 8, 53 and this study), where homeostasis is dependent upon the presence of crossover interference (refs 8, 24 and this study). In contrast to Zip2/3 foci, total crossovers show much weaker interference⁸.

Third, our beam-film model can accurately explain total crossover patterns (including coefficient of coincidence relationships and the event distribution for total crossovers) by assuming that Zip2/3 foci mark the sites of patterned (class I) crossovers; that class II crossovers represent ~30% of total crossovers; and, furthermore, that class II crossovers arise from the interactions that are 'leftover' after the operation of crossover designation and interference⁸. These 'leftover' interactions are usually matured without exchange of flanking markers; that is, to 'non-crossover' products. However, as proposed in ref. 15 and modelled in our analysis, these interactions may sometimes proceed to a crossover outcome instead of a non-crossover outcome, thus giving class II crossovers. Such a mixture of non-crossovers and a few crossovers would make the outcome for leftover meiotic interactions similar to the outcome of mitotic DSB repair.

We also note that the term 'non-interfering' is misleading when applied to class II recombinational interactions. In budding yeast, as in several (possibly all) other organisms, total recombinational interactions tend to be evenly spaced along each bivalent⁸. As a result, not only will patterned/class I crossovers exhibit interference, so too will total interactions and class II crossovers; moreover, class II crossovers will interfere with patterned (class I) crossovers⁸.

Fourth, both Zip2 and Zip3 foci occur specifically on the association sites between homologues in *zip1Δ* chromosomes^{18,50}. Analysis of Zip2 foci reveals that they exhibit interference^{8,18}. Moreover, they exhibit the same level of interference along *zip1Δ* chromosomes as along WT chromosomes when the metric of interference is physical distance (Fig. 5).

We note that this robust cytological interference contrasts with the fact that, by genetic analysis, crossover interference is significantly compromised in a *zip1Δ* mutant (see, for example, refs 44, 54). It also can be noted that cytological and genetic studies were performed in different strain backgrounds (BR at 30 °C and SK1 at 30 °C, respectively). This is because (1) in BR at 30 °C, *zip1Δ* chromosomes are well formed to permit cytological analysis but meiosis arrests during prophase, thus precluding genetic analysis of recombination outcomes, whereas (2) in SK1 at 30 °C, *zip1Δ* chromosomes are less well formed, thus making cytological analysis more difficult, whereas meiosis does not arrest, thus permitting genetic analysis.

One possible explanation for the absence of genetic interference in the *zip1Δ* mutant can be excluded. In principle, crossover designation and interference might occur normally and then be followed by a crossover-specific 'maturation defect'; that is, a defect in the probability that designated interactions will actually mature to detectable crossovers. This progression is not acceptable because, in such a situation, the detectable crossovers that do manage to form will still exhibit normal interference⁸. By contrast, a diagnostic maturation effect can be seen in an *mlh1Δ* mutant^{8,55}.

Two other, not mutually exclusive, explanations for absence of genetic interference in *zip1Δ* can be suggested, as follows.

- In WT meiosis, crossover interference is fundamentally a structure-based process to which DNA events are biochemically coupled as a downstream consequence. By this view, Zip1 would not be required for local 'crossover designation' and interference at the structural level but would be required either (1) to set up coupling between crossover/non-crossover decisions and biochemical events and/or (2) to transduce the structural interference signal into the appropriate biochemical outcome. It appears that crossover designation is a specifically programmed outcome and interactions that are not crossover-designated mature instead to non-crossovers as the default option^{15,24}. It further appears that some of these 'non-crossover-fated' interactions may actually mature into crossover products, thus giving the 'non-patterned' crossovers that are not marked by Zip3 foci⁸. Thus, in possibility (1), all interactions might progress to the 'non-crossover' outcome, giving an increase in non-crossovers and some crossovers as well, with those crossovers exhibiting the same distribution as total precursor interactions. This is, in fact, the phenotype observed at the *HIS4LEU2* hot spot in SK1 *zip1Δ* at 30 °C¹⁴. In possibility (2), crossover/non-crossover differentiation would occur at the biochemical level but there would be no progression of crossover-fated interactions. This is, in fact, the phenotype observed at the *HIS4LEU2* hot spot in SK1 *zip1Δ* at 33 °C (ref. 14).
- A reduction in the frequency of mature patterned (class I) crossovers might be accompanied by an increase in the frequency of crossovers from other sources, for example occurrence of additional DSBs, some of which then give rise to crossovers⁵⁶. Attempts to model this situation with beam-film simulations suggest that the level of extra events required to confer the strong defect in crossover interference observed in *zip1Δ* is very high (L.Z., unpublished observations). Thus, this effect may contribute to, but not be the sole basis for, absence of crossover interference in *zip1Δ*.

Fifth, localization of Zip3 along yeast chromosomes has been evaluated molecularly by chromatin immunoprecipitation (ChIP) analysis⁵⁷. This analysis identifies peaks and valleys of Zip3 abundance, genome wide, at different times of meiosis, and relates the positions of those peaks to peaks of Rec8 and Red1 (markers for chromosome axes at mid-prophase) and to peaks corresponding to DSB sites (marked by single-stranded (ss)DNA in a *dmc1Δ* strain). Zip3 is initially most prominent at centromere regions. This localization, which corresponds to the early leptotene Zip1 centromere association seen cytologically, is independent of DSB formation; it is prominent at $t = 3$ h, about the time of DSB formation; and it mostly disappears by $t = 5$ h, the time of pachytene when Zip3 foci are assayed here. Correspondingly, we find no tendency for Zip3 foci to occur at centromeres in pachytene (L.Z., unpublished observations). At $t = 4$ and 5 h, Zip3 appears in co-localization with chromosome axis markers and DNA DSB sites. Axis-localization slightly precedes DSB site localization and remains high while DSB site localization increases prominently, apparently in correlation with post-crossover-designation crossover-specific events. It is very difficult to make any relationship between ChIP results and cytological focus analysis for several reasons. (1) ChIP analysis looks at a population average localization, not a per-nucleus localization. (2) At $t = 4$ h, most cells are in leptotene/zygotene, which we do not examine cytologically. Moreover, even at $t = 5$ h, only ~50% of cells are in pachytene. Thus, ChIP data include significant signals from irrelevant stages. (3) The resolution of ChIP analysis is ~1–5 kb, with axis-association sites tending to alternate with DSB sites at separations of 5–10 kb (refs 11, 57). In contrast, Zip3 foci extend ~300 nm along the chromosome ($0.3 \pm 0.06 \mu\text{m}$; $n = 320$), which corresponds to ~90 kb in the present study (average for chromosomes III, IV and XV). Thus, a single Zip3 focus can encompass multiple axis association and DNA DSB sites. Correspondingly, ChIP analysis may well be detecting sub-focus level alterations within a crossover-designated region that reflect changes in the intimate molecular crosslinkability of Zip3 molecules to different types of DNA segment without any change in the position of the associated Zip3 focus. For example, the finding of more prominent ChIP localization to DSB sites in mutants that progress farther into recombination may reflect the extent to which those sequences are no longer buried within earlier recombination complexes. (4) To complicate matters further, it is clear cytologically that a low level of Zip3 localizes all along pachytene chromosome axes beyond that present in prominent foci. This general background will be detected in ChIP analysis but not by Zip3 focus analysis.

Visualization and definition of synaptonemal complex lengths and Zip3 focus positions (additional details in ref. 8). Meiotic time courses and sample preparation. Appropriately pre-grown cell cultures were taken through synchronous meiosis by the SPS method^{58,59}, with meiosis initiated by transfer of cells to sporulation medium ($t = 0$). Cells were harvested at $t \approx 4\text{--}5\text{ h}$, the time at which pachytene cells are most abundant (constituting approximately 50% of all cells). Harvested cells were spheroplasted to remove the cell wall and then re-suspended in MES wash (1 M sorbitol, 0.1 M MES, 1 mM EDTA, 0.5 mM MgCl_2 pH 6.5). Cells were then lysed and spread on a glass microscope slide with 1% Liposol (LIP) and fixed by 3% w/v paraformaldehyde with 3.4% w/v sucrose as described in ref. 60.

Fluorescence visualization. Glass slides with spread nuclei were incubated at room temperature for 15 min in $1\times$ Tris buffered saline (TBS) buffer (25 mM Tris-Cl, pH 8, 136 mM NaCl, 3 mM KCl) then blocked with $1\times$ TBS buffer with 1% w/v bovine serum albumin (BSA) for 10 min. Chromosomes in spread nuclei were then stained with appropriate antibodies. Primary antibodies were mouse monoclonal anti-myc (for detection of Zip3-Myc), goat polyclonal anti-Zip1 (Santa Cruz) and rabbit polyclonal anti-GFP, diluted 1:1,000 in $1\times$ TBS with 1% BSA. Secondary antibodies were anti-mouse, anti-goat and anti-rabbit IgG labelled with Alexa Fluor 488, 594 or 555 (Molecular Probes), respectively; all were diluted 1:1,000 in $1\times$ TBS with 1% BSA. Slides were mounted in Prolong Gold antifade (Molecular Probes). For condensin mutants and *spo11* hypomorphs with very low DSB levels, Zip1 staining was less bright than in WT, so axes were usually visualized by immunostaining of Rec8-3HA with rat anti-HA primary antibody and anti-rat labelled with Alexa Fluor 647 or 594 secondary antibody. Control experiments confirmed that the same synaptonemal complex lengths and Zip3 focus numbers/distributions/coefficient of coincidence relationships were obtained with either Zip1 or Rec8 staining. Stained chromosome spreads were visualized on an Axioplan IEmot microscope (Zeiss) using appropriate filters. Images were collected using Metamorph (Molecular Devices) image acquisition.

Defining Zip3 focus positions and synaptonemal complex lengths. Images for Zip3, Zip1 (or Rec8) and LacO/LacI-GFP staining (text Fig. 1a, b) were merged and aligned. The GFP-marked chromosome was analysed in nuclei where it was unambiguously separated from other chromosomes. The segmented line-tracing tool of Image J software (National Institutes of Health) was used. Each trace was initiated at the centre of the GFP focus, which typically falls beyond the end of the synaptonemal complex (white line in Fig. 1b). The trace was continued following the path of the Zip1 (Rec8) signal for the entire length of the chromosome. As the trace encountered a position judged (by eye) to be the centre of a Zip3 focus, that position was annotated using the 'mark position' function (control M). By application of the 'zoom' function, the annotated position of each Zip3 focus could be defined at the one-pixel level ($\sim 0.067\text{ }\mu\text{m}$ under our microscope). The distal end of the Zip1 (Rec8) signal was also annotated. Synaptonemal complex length was given by the annotated position mark at the end of the trace. Importantly, by this approach, each Zip3 focus (and the value for total synaptonemal complex length) was subject to its own positioning error (evaluated below) with no accumulation of error along the trace.

Accuracy of Zip3 focus (synaptonemal complex length) positions. The accuracy of the results obtained by the above approach was evaluated in several ways. (1) Coefficient of coincidence curves are highly reproducible in multiple experiments of the same strain, as shown by the correspondence of coefficient of coincidence values among different chromosomes (Fig. 1d) and for four independent analyses of a single chromosome⁸. (2) The intensity of Zip3 can be determined quantitatively along the trace and the positions of intensity peaks compared with the positions of foci defined by eye. The two methods give virtually identical results except that the eye can distinguish a significant number ($\sim 5\%$) of foci that are not, or less, obvious in the trace (for example, as shoulders on major peaks). (3) To determine the precision with which each focus position (or each synaptonemal complex length) is defined in a given trace, chromosome XV was traced six times in each of four nuclei. The four bivalents exhibited four Zip3 foci (one case) or five Zip3 foci (three cases). The variation in the absolute position of a given focus (or synaptonemal complex length) among a set of six duplicate traces ranged from 0 to $0.14\text{ }\mu\text{m}$ with an average of $0.08\text{ }\mu\text{m}$ (80 nm). Furthermore, for each focus among six traces, the standard deviation of this variation ranged from 0.02 to $0.04\text{ }\mu\text{m}$. In summary, the absolute position of each Zip3 focus (or total synaptonemal complex length) for a given traced bivalent is specified with an accuracy of approximately one pixel (67 nm).

We also performed reconstruction experiments to assess the possible effects of one-pixel accuracy on coefficient of coincidence curves. For four WT and two *pCLB2-TOP2* experimental data sets, independently, Zip3 focus positions were subjected to computational 'adjustment', with the position of each focus moved by one pixel in one direction or the other, randomly for different foci. The coefficient of coincidence curve was then re-calculated. The values of L_{CoC} were

not changed ($0.3 \pm 0.01\text{ }\mu\text{m}$ before and after 'adjustment'; further discussion of the accuracy of the coefficient of coincidence curves below). There were very subtle changes in the shape of the coefficient of coincidence curve. However, the nature of these changes in fact suggests that the relationships from the position-randomized data set represent a degradation of the more robust interference relationships observed in the primary data. (1) At smaller inter-interval distances ($< 0.2\text{ }\mu\text{m}$), coefficient of coincidence values are slightly higher. This is expected by the fact that randomized movement will artificially increase the fraction of closer-together focus pairs. (2) At larger inter-interval distances, coefficient of coincidence values fail to rise above one. This is expected because randomized movement will reduce the tendency for the inter-focus position to exhibit a node at the most likely inter-crossover position(s) (further explanation in next section).

Analysis of Zip3 focus (crossover) patterns: coefficient of coincidence and modified coefficient of coincidence relationships. Coefficient of coincidence relationships (see, for example, Fig. 1d). The coefficient of coincidence analysis is the classic indicator of crossover interference⁶¹. If done correctly (with a sufficiently large number of intervals) with a sufficiently large data set, coefficient of coincidence curves provide a highly accurate description of crossover patterns (discussion in ref. 8). We note that, in contrast, mathematical analysis of 'evenness' by application of the gamma distribution, while 'model-independent', can give a misleading impression with respect to mutant phenotypes or other types of variation (discussion in ref. 8). For example, a defect in maturation of crossovers after their positions have been designated has no effect on interference and thus does not affect coefficient of coincidence relationships but significantly alters the value of the gamma 'evenness' parameter. Coefficient of coincidence curves for Zip3 foci were obtained using the 'Analyze crossover data' feature of the beam-film program, using as an input the experimentally defined positions of Zip3 foci in a given experiment⁸. For this purpose, chromosomes are divided into a number of intervals with equal size (detailed discussions in ref. 8 protocol S1). For each interval the total frequency of Zip3 foci in the set of chromosomes examined is determined. Then, for each pair of intervals, the observed frequency of chromosomes exhibiting a Zip3 focus in both intervals (referred to for convenience as 'double crossovers') is determined. This value defines the frequency of 'observed double crossovers'. If crossovers (Zip3 foci) arise independently in each interval, the predicted frequency of double crossovers for a given pair of intervals should be the product of the frequencies of crossovers (Zip3 foci) in the two intervals considered individually. This product is the frequency of 'expected double crossovers'. The coefficient of coincidence for that particular pair of intervals is the ratio of these two frequencies, that is the observed/expected ratio for that interval pair. A coefficient of coincidence curve is obtained by considering all possible pairs of intervals, with the coefficient of coincidence value for each pair plotted as a function of the distance between (the midpoints of) the two corresponding intervals. For a classic coefficient of coincidence curve, at very small inter-interval distance, the coefficient of coincidence is close to zero, indicating very strong crossover interference. As the inter-interval distance increases, the coefficient of coincidence also gradually increases, indicating that crossover interference decreases with increased inter-interval distance. Eventually, the coefficient of coincidence value reaches one, implying that, at the corresponding inter-interval distance, crossover interference no longer has any influence. At certain specific larger inter-interval distances, the coefficient of coincidence value tends to be greater than one, implying that, at these distances, there is a higher probability of double crossovers than predicted on the basis of independent occurrence. Nodes of coefficient of coincidence greater than 1 tend to occur at inter-interval distances that correspond approximately to the average inter-crossover distance and multiples thereof (see ref. 8 for more examples). This pattern reflects the fact that operation of crossover interference tends to create an evenly spaced array of crossovers (Zip3 foci, in this analysis).

For convenience, the inter-interval distance at which the coefficient of coincidence = 0.5 is defined as L_{CoC} and can be used as a measurement for 'crossover interference strength', by which is meant the effective distance over which crossover interference acts. Importantly, at a mechanistic level, variations in L_{CoC} can result from variations in features other than the distance over which the interference signal spreads (for example, as discussed for beam-film simulations below). Values of L_{CoC} are highly reproducible from one experiment to another. For the three analysed chromosomes in WT meiosis, values for individual experiments and the averages and standard deviations are as follows: chromosome XV, 0.31, 0.3, 0.32, 0.32 (0.31 ± 0.01 ; $n = 4$); chromosome III, 0.31, 0.32, 0.3 (0.31 ± 0.01 ; $n = 3$); chromosome IV, 0.31, 0.32 (0.32 ± 0.01 ; $n = 2$). Further documentation is in ref. 8.

Modified coefficient of coincidence analysis (Fig. 1e). As an alternative approach to evaluating the effective interference distance, we adapted the 'modified coefficient of coincidence' approach previously described for analysis of genetic crossover data²¹. For the present purpose, each interval is used as a reference (Ref; Fig. 1e top left). Chromosomes are then divided into two groups: those with or without a crossover (Zip3 focus) in this reference interval (CO^+_{R} or CO^-_{R}). Another nearby

interval is then selected as a test (Test (T)). For each reference group (CO^+_R or CO^-_R), the numbers of chromosomes with and without a crossover in this test interval is determined (CO^+_T and CO^-_T). If crossover levels are lower in the CO^+_R group than in the CO^-_R group, the presence of a crossover in the reference interval has reduced the probability of a crossover in the Test interval; that is, interference emanating from the reference interval has been felt in that Test interval. When this evaluation is performed for all intervals in the vicinity of a given reference interval, it reveals the distance over which interference extends outward from that interval, giving L_{MCOC} for that reference interval (Fig. 1e, top right). Determination of L_{MCOC} values for all intervals along each of the three analysed chromosomes gives an average L_{MCOC} for that chromosome (Fig. 1e, bottom right).

This analysis requires an evaluation, for each comparison between a reference interval and a test interval, of whether the relative frequencies of CO^+_T and CO^-_T chromosomes are the same for the CO^+_R and CO^-_R groups or different (that is, lower in the CO^+_R group). For this purpose, Fisher's exact test was applied. Since interference is stronger (and thus more likely to be statistically significant) at shorter distances, the more stringent the probability specified by Fisher's exact test, the shorter the inferred 'interference distance'. The standard criterion for significance by this method is $P < 0.05$. By this criterion, L_{MCOC} for the three analysed chromosomes in WT meiosis was $0.3\ \mu\text{m}$, which is the same as L_{COC} as defined above. With a more stringent criterion, $P < 0.01$, L_{MCOC} is slightly shorter ($0.25\ \mu\text{m}$). Importantly, mutants with decreased interference distance always showed decreased L_{MCOC} compared with WT regardless of whether the standard, or more stringent, criterion was applied. Thus, when $P < 0.05$, L_{MCOC} in *top2* mutants versus WT was 1.3 intervals versus 1.9 intervals (that is, $0.2\ \mu\text{m}$ versus $0.3\ \mu\text{m}$); when $P < 0.01$, L_{MCOC} in *top2* mutants versus WT was 1.0 versus 1.5 in WT (that is, $0.16\ \mu\text{m}$ versus $0.25\ \mu\text{m}$). Given that $P < 0.05$ is the standard value applied for Fisher's exact test and the fact that L_{COC} and L_{MCOC} correspond at $P < 0.05$, we adopted this level of stringency to describe L_{MCOC} in the present analysis (Figs 1, 2 and 4 and Extended Data Fig. 3).

Beam-film simulations. The beam-film model and the program used for simulations are described in detail in refs. 4 and 8. The beam-film program was recently rewritten in MATLAB (R2010a), which is downloadable at <https://app.box.com/s/hv91q2nrtq0cp9n8iy9m>.

Outline of the beam-film model. An array of precursor interactions comes under global stress, which causes a first (most sensitive) precursor to go critical, undergoing a stress-promoted change that commits it to becoming a crossover ('crossover designation'). The intrinsic effect of this change will be a local reduction in the level of stress at the site of the change. To even out distribution of stress along the chromosome, the initial local reduction in stress then redistributes outwards in both directions, thus reducing the probability that any subsequent crossover designation(s) will occur in the affected region. This effect constitutes crossover interference. Assuming that the system does not comprise a single elastic component, the extent of stress reduction will dissipate with increasing distance away from the nucleation site, becoming negligible over a characteristic distance (corresponding to the 'interference distance'). A second crossover designation may then occur. If so, that crossover will occur preferentially at a position that retains a high stress level and thus preferentially at some distance away from the position of the prior crossover designation. This second crossover designation will again result in local stress relief and redistribution (and thus interference), giving a new stress landscape along the chromosome. If/as additional events occur, they will tend to fill in the holes between prior events, thus giving an evenly spaced array. The beam-film model predicts the number and array of crossovers that will occur in particular system with particular mechanical properties that are analogous to a known system in the physical world (the 'beam-film system'). In this particular system, the magnitude of the stress reduction decreases exponentially with distance away from its nucleation point.

Beam-film best-fit simulations. In beam-film simulation analysis, the parameters of the beam-film model are varied to define the constellation of parameter values at which the predicted array of crossover events best matches that observed experimentally for a particular data set⁸. As described in detail elsewhere⁸, the parameters to be specified fall into three categories that describe, respectively, the following: (1) the array of precursor interactions upon which crossover patterning acts; (2) the nature of the patterning process per se; and (3) the probability that a crossover-designated interaction will actually mature to an experimentally detectable crossover or crossover marker (that is, a Zip3 focus).

For modelling, the level of global stress is progressively increased up to a maximum specified level (S_{max}). As the level of stress increases, precursors will undergo crossover designation sequentially in relation to their relative local stress levels at that moment in the sequence of events (differently for different bivalents according to their specific histories). Each crossover designation triggers reduction in stress, in both directions, over a characteristic length given by a specific parameter (L). The value of L for a particular simulation is directly reflected in the

resultant coefficient of coincidence relationships and corresponds very closely to the inter-interval distance at which the coefficient of coincidence = 0.5, defined here as L_{BF} . A third patterning parameter ('A') describes precursor reactivity: that is, the way in which the probability of crossover designation varies as a function of the local stress level at the corresponding position. A fourth patterning parameter ('clamping') permits adjustment of crossover probabilities near chromosome ends.

Parameter values for beam-film best-fit simulations of crossovers (Zip3 foci) along WT yeast chromosomes are described in ref. 8. The best-fit simulations for mutant patterns presented in Figs 2a, 3a–c and 4a, b (except mutants with altered axis lengths) were obtained using these same parameter values except that the value of L was appropriately reduced, from $\sim 0.3\ \mu\text{m}$ to $\sim 0.2\ \mu\text{m}$, resulting in a commensurate reduction in L_{BF} . Best-fit simulations in situations with altered DSB levels (Fig. 2d) also involved changes in the number of precursors (N), as discussed below ('crossover homeostasis analysis') and in Extended Data Fig. 4. Best-fit simulations in mutants with altered axis lengths also involved changes in the number of precursors (N), as discussed in Extended Data Fig. 9.

Crossover homeostasis analysis. Crossover homeostasis is a nonlinear relationship between the number of DSBs and the number of crossovers^{8,24}. The existence and magnitude of crossover homeostasis depends on the existence and strength of crossover interference (see text and ref. 8).

Beam-film simulations of crossover homeostasis. A beam-film best-fit simulation predicts the number of crossovers that will occur if crossover designation and interference occur according to a specific set of values for involved parameters. To get a simulated crossover homeostasis curve under a particular set of conditions, multiple beam-film simulations were performed at different values of the precursor number N , which were varied over a desired range, and with the values of all other parameters held constant. The average numbers of crossovers predicted for each evaluated value of N were then plotted as a function of N . Such curves were then obtained analogously at different values for the interference distance L (ref. 8; Fig. 2d).

Experimental evaluation of crossover homeostasis by Zip3 focus analysis. The positions of Zip3 foci were determined along specific marked chromosomes (XV and III) in a series of strain backgrounds known to give varying levels of DSBs, in both a *TOP2* and a *pCLB2-TOP2* background. Coefficient of coincidence relationships and the numbers and distributions of Zip3 foci per bivalent for all strains are given in Figs 1 and 2 and Extended Data Figs 2 and 4. Average Zip3 focus numbers per chromosome (average \pm s.d.) are shown in Fig. 2d and listed in the legend to Extended Data Fig. 4.

DSB levels were decreased below WT levels by a previously described series of hypomorphic *spo11* alleles (*spo11HA*, *spo11YFHA*, *spo11DAHA*; ref. 24). DSB levels were increased above WT levels using a *tel1A* mutation, alone and in combination with a *spo11* hypomorph (*tel1A spo11HA*). The average numbers of Zip3 foci per bivalent in the different strains were then plotted as a function of beam-film precursor or DSB level (discussion below). Such analysis was in strain backgrounds that were also either (1) WT for crossover interference (*TOP2*) or (2) carried the *pCLB2-TOP2* construct that resulted in meiotic depletion of topoisomerase II (see text).

The number of DSBs per bivalent in a *TOP2* strain with WT DSB formation can be accurately determined on the basis of comprehensive evaluation results from DSB mapping (for example, ref. 12), microarray (for example, ref. 45) and classic genetic measurements (<http://www.yeastgenome.org>). The numbers of DSBs on chromosomes III, IV and XV are thus defined as 6, 19 and 13 respectively. The relative levels of DSBs in strains carrying *spo11* mutations has been evaluated in a *TOP2* background by gel electrophoresis in a *rad50S* background²⁴ (where DSBs do not turn over). In the *tel1A* mutant, DSBs are increased by $\sim 50\%$ at the *HIS4LEU2* locus in a *rad50S* background without significantly altering crossover interference^{8,62} (Extended Data Fig. 7 and L.Z., unpublished observations).

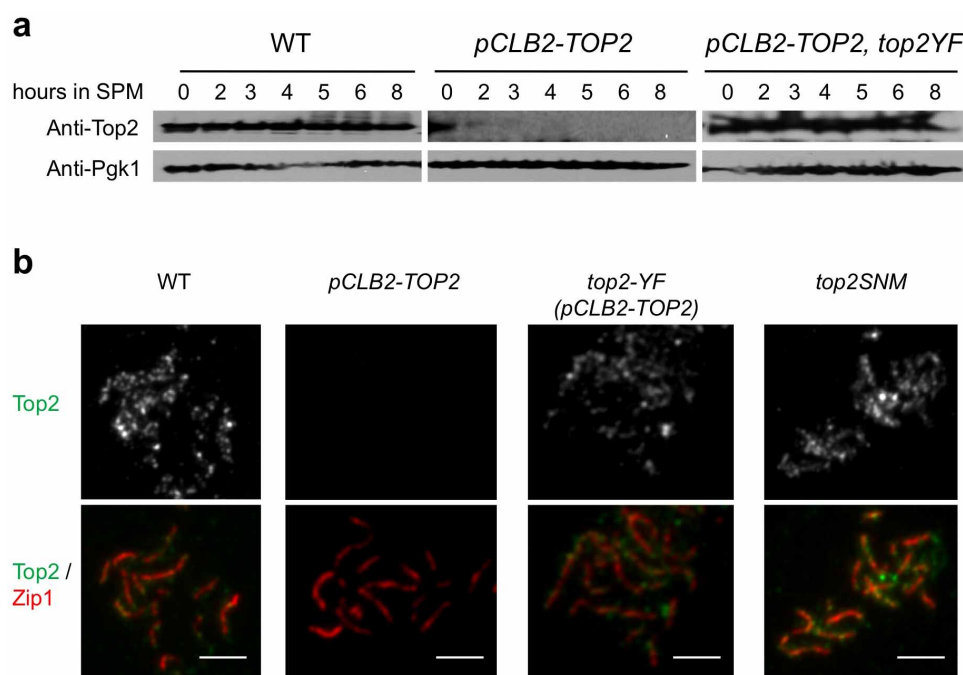
However, in some regions and circumstances, *rad50S* DSB levels are known to be lower than the level of DSBs in *RAD50* meiosis (see, for example, refs 11, 12). Furthermore, *rad50S* analysis of *spo11/tel1A* alleles in a *pCLB2-TOP2* background has not been performed. We therefore also evaluated DSB levels by application of beam-film analysis. For all strains analysed for Zip3 focus patterns, both *TOP2* and *pCLB2-TOP2*, best-fit beam-film simulations were defined⁸ (Figs 2–4 and Extended Data Figs 2 and 4). For each strain, all parameter values were held constant at those defined for the two *SPO11 TEL1* cases (see text) except that the average number of precursors per bivalent (N) was varied to determine the value that gave the optimal match between observed and predicted crossover patterns for that strain. Beam-film-predicted DSB/precursor levels were the same for the *TOP2* and *pCLB2-TOP2* versions of all strains (Figs 2–4 and Extended Data Fig. 4c). This prediction matches the experimental finding that *TOP2* and *pCLB2-TOP2* strains exhibit the same level of total inter-homologue events (crossover plus non-crossover) at *HIS4LEU2* in a *RAD50 SPO11 TEL1* background (Extended Data Fig. 8). Furthermore, for *TOP2* strains, DSB/precursor

values obtained by beam-film simulations are very similar to those obtained on the basis of *rad50S* analysis (Extended Data Fig. 4c). Correspondingly, crossover homeostasis relationships are very similar regardless of whether DSBs or beam-film-predicted precursors are used as the metric (Fig. 2d and Extended Data Fig. 4d).

Interestingly, experimentally determined *rad50S* DSB levels tend to be slightly lower than those predicted by beam-film analysis, especially at lower DSB levels (Extended Data Fig. 4). Moreover, experimental data match beam-film-predicted crossover homeostasis relationships somewhat more accurately when the metric of the DSB level is the beam-film-predicted precursor level, especially at lower DSB/precursor levels (Extended Data Fig. 4d). This correspondence suggests that beam-film-predicted values may be more accurate than *rad50S* experimental values. Data in ref. 24 support this conclusion: at *HIS4LEU2*, a *spo11HA/HA* strain exhibits 50% of the *SPO11* level of *rad50S* DSBs but 62% of the level of inter-homologue recombination products (crossover plus non-crossover), implying a deficit of 20% by *rad50S* analysis. Similarly, a *spo11HA/DA* strain exhibits 20% of the *SPO11* level of *rad50S* DSBs but 27% the level of inter-homologue recombination products, a deficit of 26%.

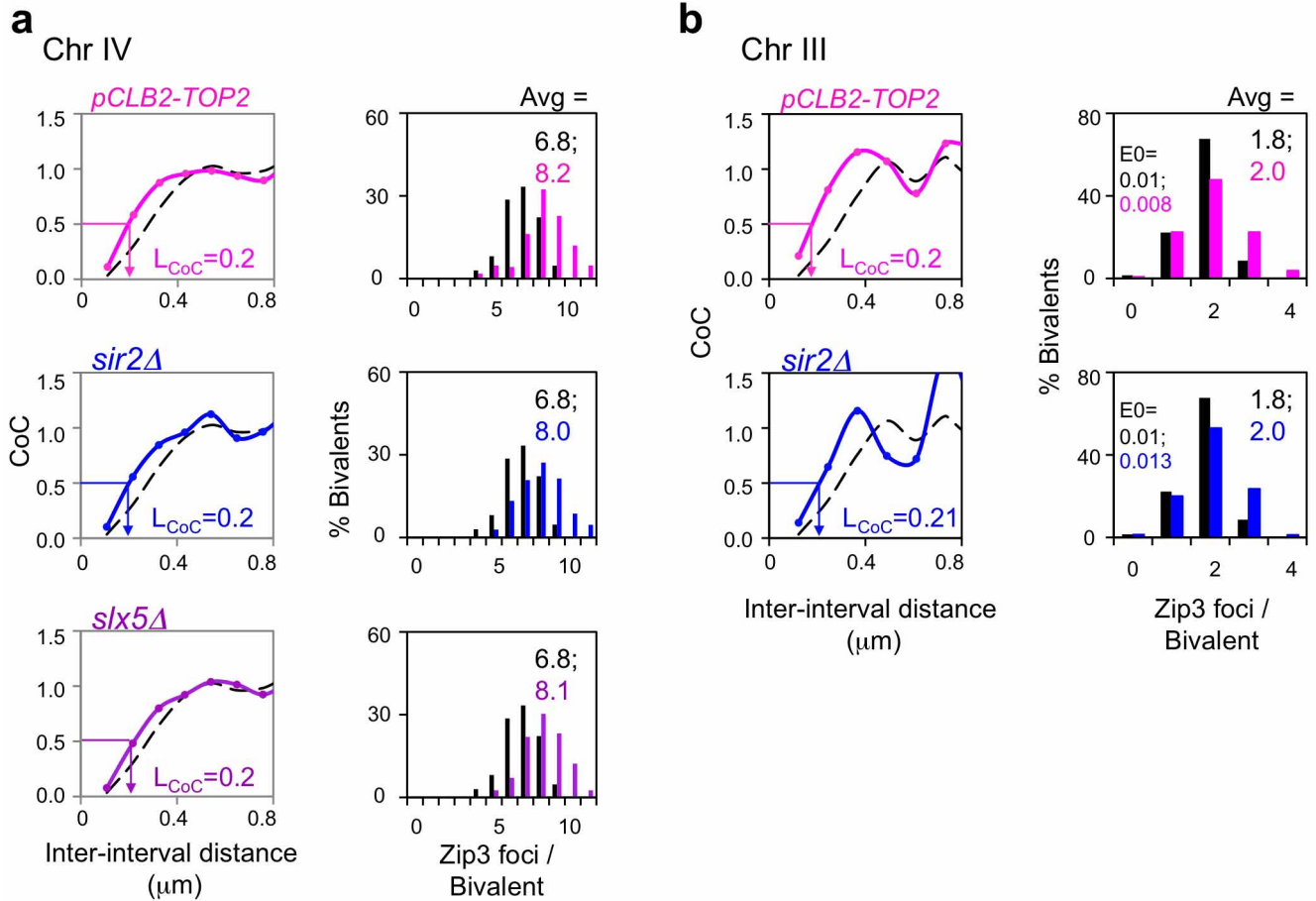
These analyses also provide further evidence (in addition to that presented in Extended Data Fig. 7) that the increased number of Zip3 foci seen in *top2* mutants compared with *TOP2* strains cannot be explained as increased DSBs.

44. Chen, S. Y. *et al.* Global analysis of the meiotic crossover landscape. *Dev. Cell* **15**, 401–415 (2008).
45. Mancera, E., Bourgon, R., Brozzi, A., Huber, W. & Steinmetz, L. M. High-resolution mapping of meiotic crossovers and non-crossovers in yeast. *Cell* **454**, 479–485 (2008).
46. Cherry, J. M. *et al.* Genetic and physical maps of *Saccharomyces cerevisiae*. *Nature* **387**, 67–73 (1997).
47. Argueso, J. L., Wanat, J., Gemici, Z. & Alani, E. Competing crossover pathways act during meiosis in *Saccharomyces cerevisiae*. *Genetics* **168**, 1805–1816 (2004).
48. de los Santos, T. *et al.* The Mus81/Mms4 endonuclease acts independently of double-Holliday junction resolution to promote a distinct subset of crossovers during meiosis in budding yeast. *Genetics* **164**, 81–94 (2003).
49. Hollingsworth, N. M., Ponte, L. & Halsey, C. MSH5, a novel MutS homolog, facilitates meiotic reciprocal recombination between homologs in *Saccharomyces cerevisiae* but not mismatch repair. *Genes Dev.* **9**, 1728–1739 (1995).
50. Chua, P. R. & Roeder, G. S. Zip2, a meiosis-specific protein required for the initiation of chromosome synapsis. *Cell* **93**, 349–359 (1998).
51. Shinohara, M., Oh, S. D., Hunter, N. & Shinohara, A. Crossover assurance and crossover interference are distinctly regulated by the ZMM proteins during yeast meiosis. *Nature Genet.* **40**, 299–309 (2008).
52. Jessop, L., Rockmill, B., Roeder, G. S. & Lichten, M. Meiotic chromosome synapsis-promoting proteins antagonize the anti-crossover activity of Sgs1. *PLoS Genet.* **2**, e155 (2006).
53. Henderson, K. A. & Keeney, S. Tying synaptonemal complex initiation to the formation and programmed repair of DNA double-strand breaks. *Proc. Natl Acad. Sci. USA* **101**, 4519–4524 (2004).
54. Sym, M. & Roeder, G. S. Crossover interference is abolished in the absence of a synaptonemal complex protein. *Cell* **79**, 283–292 (1994).
55. Nishant, K. T. *et al.* The baker's yeast diploid genome is remarkably stable in vegetative growth and meiosis. *PLoS Genet.* **6**, e1001109 (2010).
56. Kauppi, L. *et al.* Numerical constraints and feedback control of double-strand breaks in mouse meiosis. *Genes Dev.* **27**, 873–886 (2013).
57. Serrentino, M. E., Chaplais, E., Sommermeyer, V. & Borde, V. Differential association of the conserved SUMO ligase Zip3 with meiotic double-strand break sites reveals regional variations in the outcome of meiotic recombination. *PLoS Genet.* **9**, e1003416 (2013).
58. Kim, K. P. *et al.* Sister cohesion and structural axis components mediate homolog bias of meiotic recombination. *Cell* **143**, 924–937 (2010).
59. Koszul, R. & Kleckner, N. Dynamic chromosome movements during meiosis: a way to eliminate unwanted connections? *Trends Cell Biol.* **19**, 716–724 (2009).
60. Loidl, J., Klein, F. & Engebrecht, J. Genetic and morphological approaches for the analysis of meiotic chromosomes in yeast. *Methods Cell Biol.* **53**, 257–285 (1998).
61. Charles, D. R. The spatial distribution of cross-overs in X-chromosome tetrads of *Drosophila melanogaster*. *J. Genet.* **36**, 103–126 (1938).
62. Zhang, L., Kim, K. P., Kleckner, N. E. & Storlazzi, A. Meiotic double-strand breaks occur once per pair of (sister) chromatids and, via Mec1/ATR and Tel1/ATM, once per quartet of chromatids. *Proc. Natl Acad. Sci. USA* **108**, 20036–20041 (2011).
63. Imai, S., Armstrong, C. M., Kaerberlein, M. & Guarente, L. Transcriptional silencing and longevity protein Sir2 is an NAD-dependent histone deacetylase. *Nature* **403**, 795–800 (2000).
64. Wu, C. S., Chen, Y. F. & Gartenberg, M. R. Targeted sister chromatid cohesion by Sir2. *PLoS Genet.* **7**, e1002000 (2011).
65. Dhillion, N. & Kamakaka, R. T. A histone variant, Htz1p, and a Sir1p-like protein, Esc2p, mediate silencing at HMR. *Mol. Cell* **6**, 769–780 (2000).
66. Derbyshire, M. K., Weinstock, K. G. & Strathern, J. N. HST1, a new member of the SIR2 family of genes. *Yeast* **12**, 631–640 (1996).
67. Hong, S. *et al.* The logic and mechanism of homologous recombination partner choice. *Mol. Cell* **51**, 440–453 (2013).



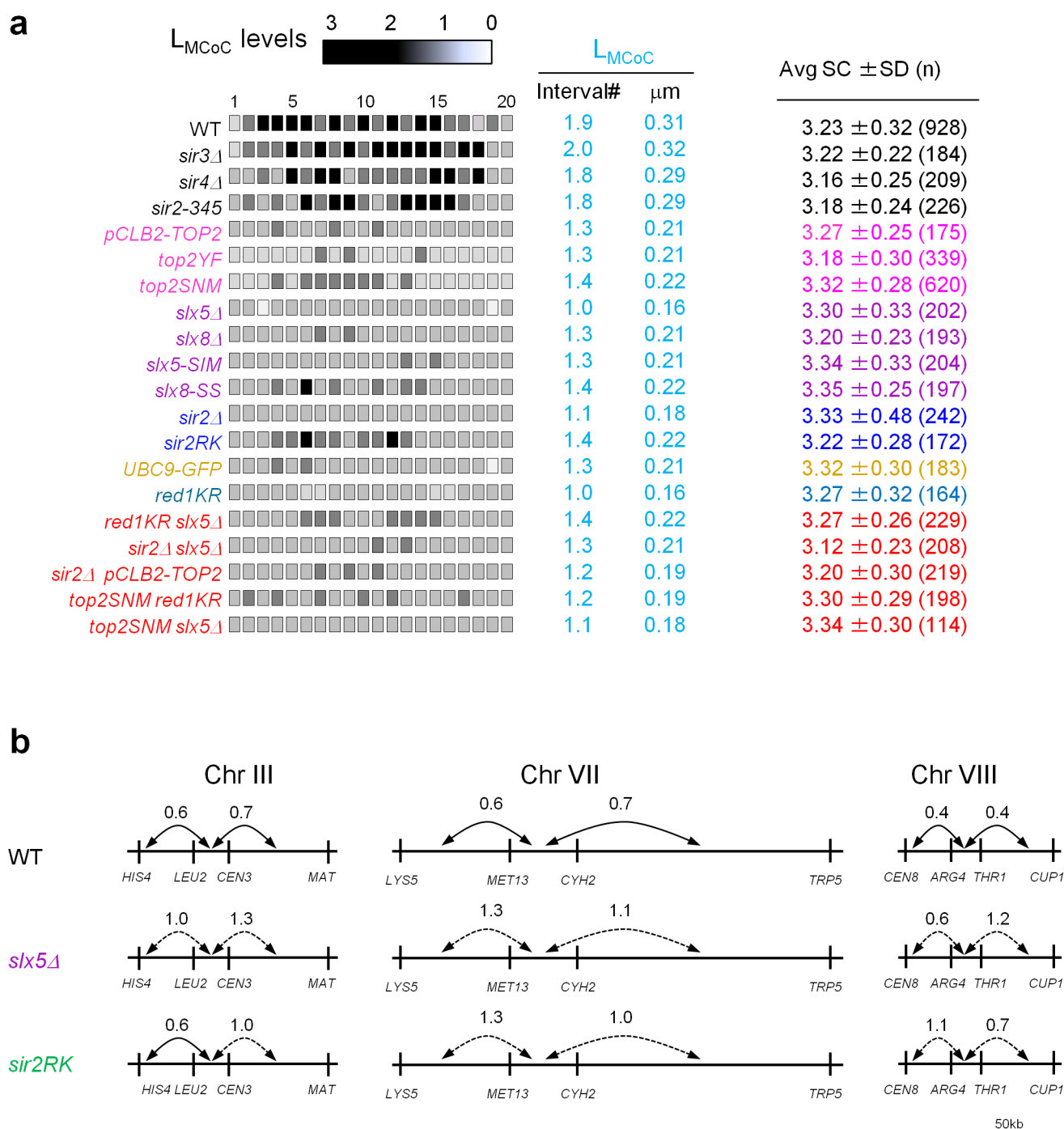
Extended Data Figure 1 | Top2 protein level and localization on chromosomes in three *top2* mutants. **a**, Top2 protein levels shown as a function of time after entry into meiosis ($t = 0$). Top2 levels are severely reduced in *pCLB2-TOP2* (middle panel) and are the same as WT in *pCLB2-TOP2 top2YF* ($\pm 20\%$ relative to anti-Pgk1 control). Western blot analysis used anti-Top2 antibody (TopoGEN 2014) and anti-Pgk1 antibody (Abcam

ab113687). **b**, Immunostaining of Top2 on meiotic chromosomes with the same antibody used for western blot analysis in **a**: at pachytene (shown) and at leptotene (data not shown). Top2 is undetectable on chromosomes in *pCLB2-TOP2* and is present at similar levels to WT in *pCLB2-TOP2 top2YF* and *top2SNM*. Chromosomes were concomitantly immunostained for Zip1 (Santa Cruz, sc-48716) as in text Fig. 1. Scale bars, 3 μm .



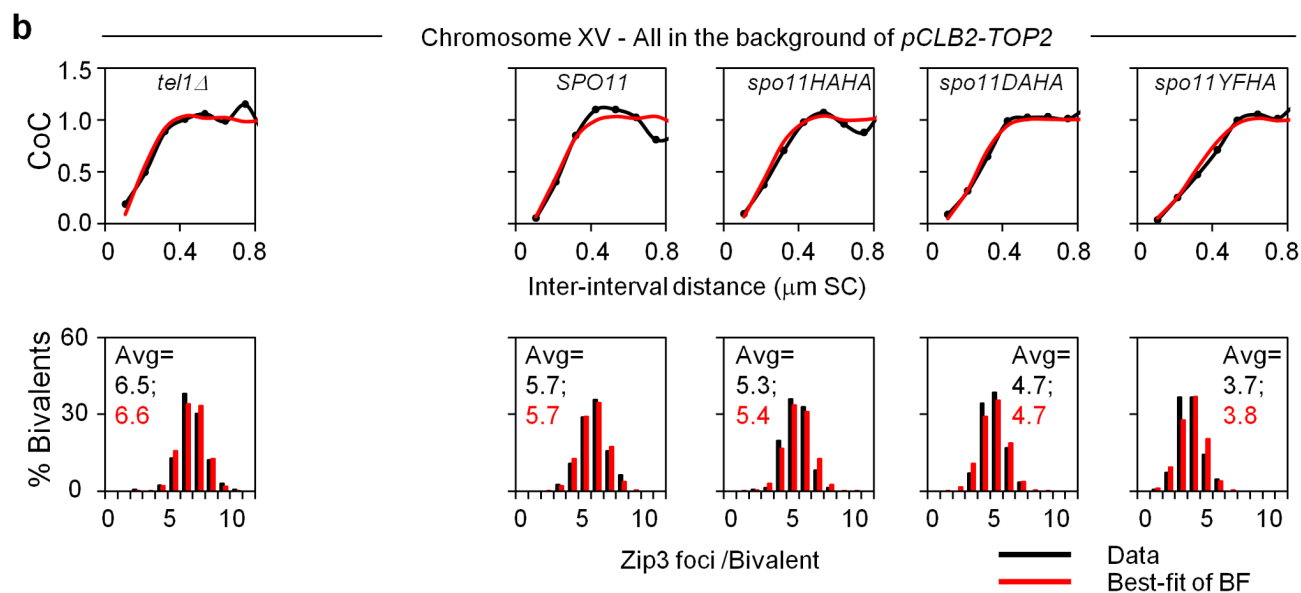
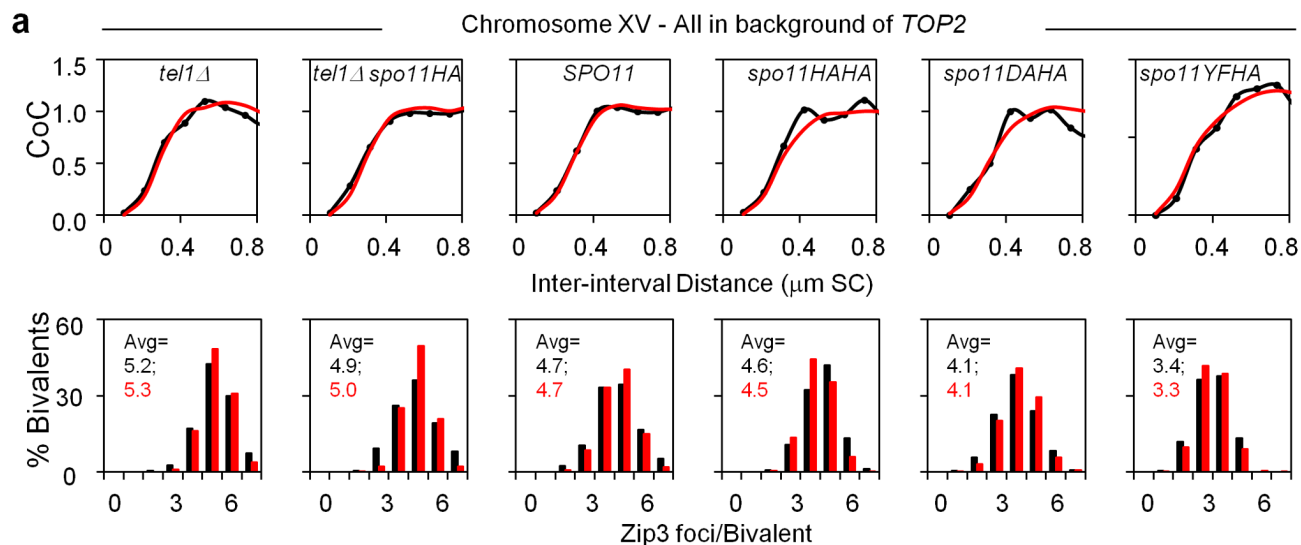
Extended Data Figure 2 | Decreased crossover interference in *pCLB2-TOP2* and *sir2 Δ* , *slx5 Δ* is confirmed on other chromosomes. a, b, The same decreases in crossover interference ($L_{\text{CoC}} \approx 0.2 \mu\text{m}$ versus $\approx 0.3 \mu\text{m}$ in WT) and corresponding increases crossover number observed for the indicated

mutants on chromosome XV (Figs 2 and 3) are also observed on chromosomes IV and III in *pCLB2-TOP2* and *sir2 Δ* , and on chromosome IV in *slx5 Δ* . Data for WT in black.



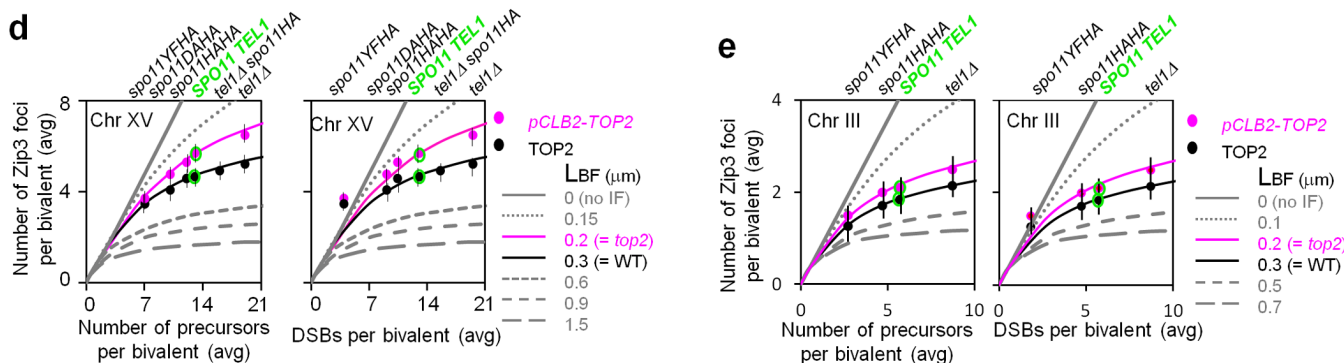
Extended Data Figure 3 | Decreased crossover interference as revealed by modified coefficient of coincidence and tetrad analysis using the method of ref. 21, but synaptonemal complex length is the same as in WT. a, By modified coefficient of coincidence analysis (Fig. 1; Methods), crossover interference can extend to about two intervals on either side of the reference interval ($L_{MCoC} \approx 0.3 \mu m$) in WT and in three *sir2* mutants that exhibit WT crossover patterning by other criteria ($L_{CoC} \approx 0.3 \mu m$; Fig. 3 and Extended Data Fig. 2). In contrast, in all examined single and double mutants where crossover interference is defective ($L_{CoC} \approx 0.2 \mu m$; Figs 2–4), crossover interference extends only about 1.3 intervals ($L_{MCoC} \approx 0.2 \mu m$) (for *top2* mutants, see also Fig. 2). Right column shows synaptonemal complex lengths for each of the analysed strains (average \pm s.d.). There is no significant difference between

strains exhibiting WT interference (average of averages is $3.25 \pm 0.06 \mu m$) and strains defective in the *top2* interference pathway (average of averages is $3.27 \pm 0.07 \mu m$). **b,** Decreased crossover interference in *slx5 Δ* and *sir2RK* as revealed by tetrad analysis. Each pair of intervals was tested, reciprocally, for the ratio of the map distances in one interval with and without crossovers in the other interval. Each number shows the average of the ratios for the two reciprocal cases. A value less than 1 indicates crossover interference. Solid and dotted lines indicate whether the level of interference is statistically ($P < 0.05$ by G-test) significant or not, respectively. Genetic crossover interference is greatly decreased in *slx5 Δ* and *sir2RK* relative to WT on each of three chromosomes. Tetrad data upon which this analysis is based are given in Supplementary Table 2.



c Comparison of *rad50S* DSB levels and BF-predicted precursor levels (N) on chromosome XV

Strains	<i>tel1Δ</i>	<i>tel1Δspo11HA</i>	<i>SPO11</i>	<i>spo11HAHA</i>	<i>spo11DAHA</i>	<i>spo11YFHA</i>
Relative DSB levels (<i>rad50S</i>)	150	120	100	80	70	30
DSBs per Chr XV	19	16	13	10.5	9.1	4
BF-predicted DSBs (precursors = (N))	19	16	13	11	10	7

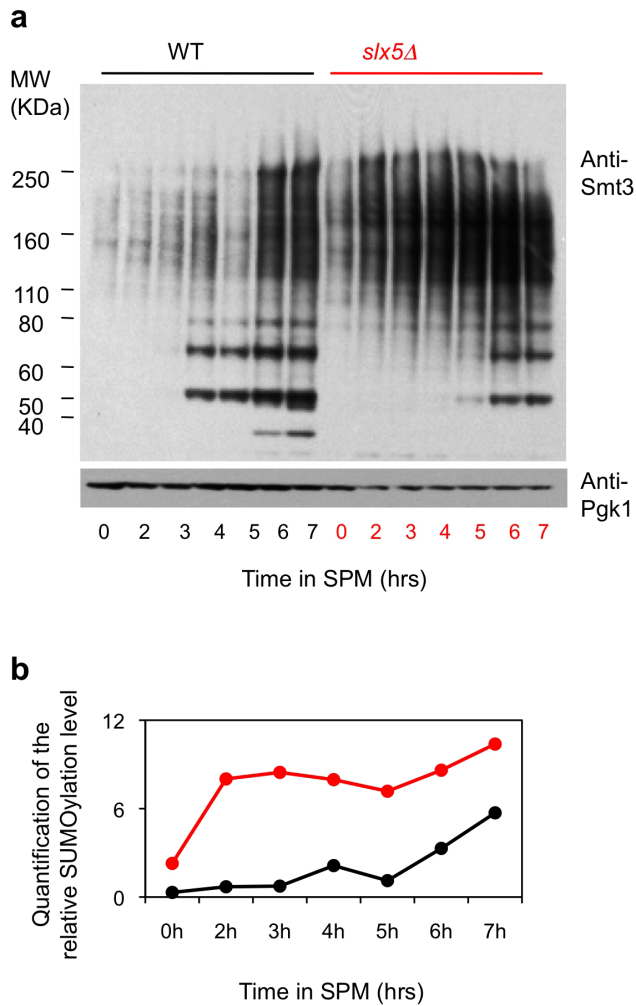


Extended Data Figure 4 | Additional aspects of crossover homeostasis analysis.

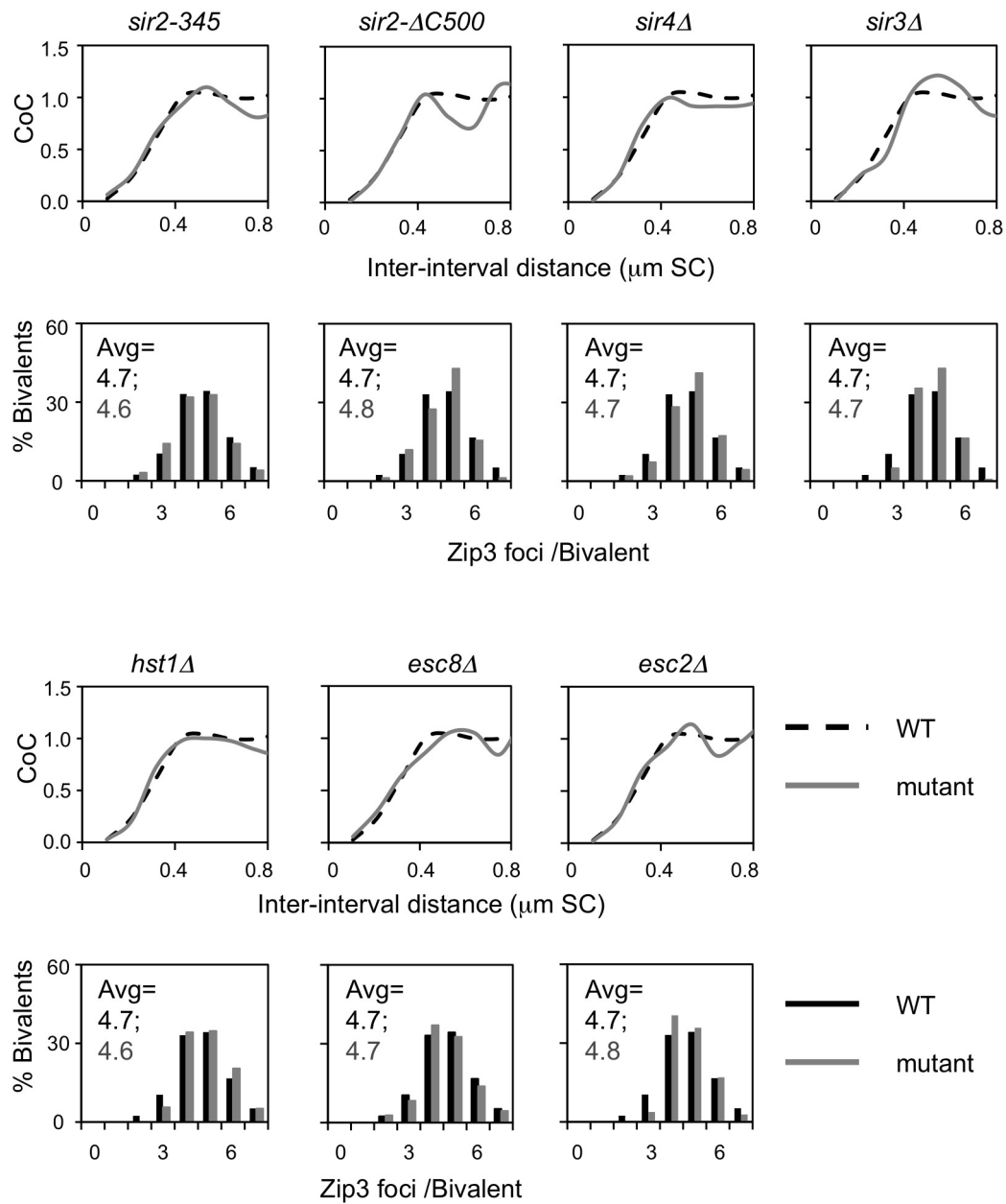
a, b, Crossover patterns along chromosome XV in *TOP2* strains (**a**, black) and *pCLB2-TOP2* strains (**b**, black) with WT or altered DSB levels as conferred by the indicated *spo11/tel1* genotypes (for crossover homeostasis analysis; Fig. 2d and Methods). All experimental data sets were also subjected to beam-film simulation analysis (**a** and **b**, red). In all cases (**a** and **b**, red), best-fit simulations were obtained by using the same parameters as those that give the best-fit for *SPO11 TEL1* meiosis (ref. 8; Fig. 2a) except that number of precursors (given by parameter *N*) was altered to account for alterations in DSB levels in the different strain backgrounds ($L_{BF} = 0.3 \mu\text{m}$ in *TOP2* background versus $0.2 \mu\text{m}$ in *pCLB2-TOP2* background; see Methods and below). For each *spo11/tel1* genotype, the best-fit value of (*N*) is the same in *pCLB2-TOP2* as in *TOP2*, thus confirming that the only change in various *pCLB2-TOP2* strains examined is a change in precursor number, with no change in interference. The same results are also seen for beam-film simulations of analogous data for chromosome III (not shown). These results further illustrate the accuracy with which beam-film simulations can describe diverse crossover patterns.

c, Comparison of *rad50S* DSB levels and beam-film-predicted precursor levels (*N*) for chromosome XV among strains with varying DSB levels due to different *SPO11 TEL1* or carrying *spo11* and/or *tel1* mutant alleles. Top line: number of DSBs genome-wide, relative to WT = 100, as defined by *rad50S* analysis in *TOP2* strains, either *SPO11 TEL1* or carrying *spo11* and/or *tel1* mutant alleles (details in Methods). Middle line: number of DSBs predicted for chromosome XV. Number of DSBs in *TOP2 SPO11 TEL1* was defined by several approaches (details in Methods). DSBs per chromosome XV as predicted for *spo11/tel1* mutant strains by comparison of *rad50S* DSB levels with *SPO11 TEL1* (top line). Bottom line: number of precursors predicted to be present by beam-film best-fit simulation analysis (given by parameter *N*, above). Predicted values are the same for *TOP2* and *pCLB2-TOP2* strain series (from simulations in **a** and **b**).

Note that in strains with lower total DSB levels, *rad50S* analysis gives lower DSB/precursor levels than beam-film simulations (discussion in Methods). Analogous results are obtained for chromosome III, as follows. (1) The predicted values of *N* are the same for both *TOP2* and *pCLB2-TOP2* strain series: *N* = 9 for *tel1A*, 6 for *TEL1 SPO11*, 5 for *spo11-HA/spo11HA* and 3 for *spo11-HA/spo11YF*. (2) These predicted values of *N* correspond well to DSB values predicted from *rad50S* analysis except at the lowest DSB levels: predicted DSBs = 9 for *tel1A*, 6 for *TEL1 SPO11*, 5 for *spo11-HA/spo11HA* and 2 for *spo11-HA/spo11YF*. **d,** Experimentally determined numbers of Zip3 foci from the analyses of chromosome XV in **a** and **b** are plotted as a function of either the number of precursors predicted by beam-film simulation analysis (left) or the number of DSBs predicted by *rad50S* DSB analysis (right) (values from **c**). **e,** Same as **d**, except that we analysed chromosome III. A slightly better match of experimental data to beam-film simulation predictions is obtained when the *x* axis metric is the predicted precursor number than when it is *rad50S* predicted DSB levels, suggesting that beam-film simulations are more accurate than *rad50S* DSB analysis, which is known to underestimate DSBs in several situations. Note that for each strain and chromosome, Zip3 foci were analysed in 200–300 cells. The average numbers of foci per bivalent \pm s.d. as presented in **d** and **e** were as follows. *TOP2* chromosome XV (**d**): *tel1A* 5.21 ± 0.93 ; *tel1A spo11HA* 4.92 ± 1.12 ; *TEL1 SPO11* 4.67 ± 1.16 ; *spo11HA/spo11HA* 4.11 ± 0.97 ; *spo11HA/spo11DA* 4.07 ± 1.07 ; *spo11HA/spo11YF* 3.51 ± 0.88 . *pCLB2-TOP2* chromosome XV (**d**): *tel1A* 6.46 ± 1.13 ; *TEL1 SPO11* 5.96 ± 1.1 ; *spo11HA/spo11HA* 5.29 ± 0.99 ; *spo11HA/spo11DA* 4.76 ± 0.94 ; *spo11HA/spo11YF* 3.71 ± 0.98 . *TOP2* chromosome III (**e**): *tel1A* 2.16 ± 0.59 ; *TEL1 SPO11* 1.82 ± 0.55 ; *spo11HA/spo11HA* 1.7 ± 0.62 ; *spo11HA/spo11YF* 1.31 ± 0.66 . *pCLB2-TOP2* chromosome III (**e**): *tel1A* 2.49 ± 0.82 ; *TEL1 SPO11* 2.1 ± 0.87 ; *spo11HA/spo11HA* 2.07 ± 0.75 ; *spo11HA/spo11YF* 1.51 ± 0.69 .

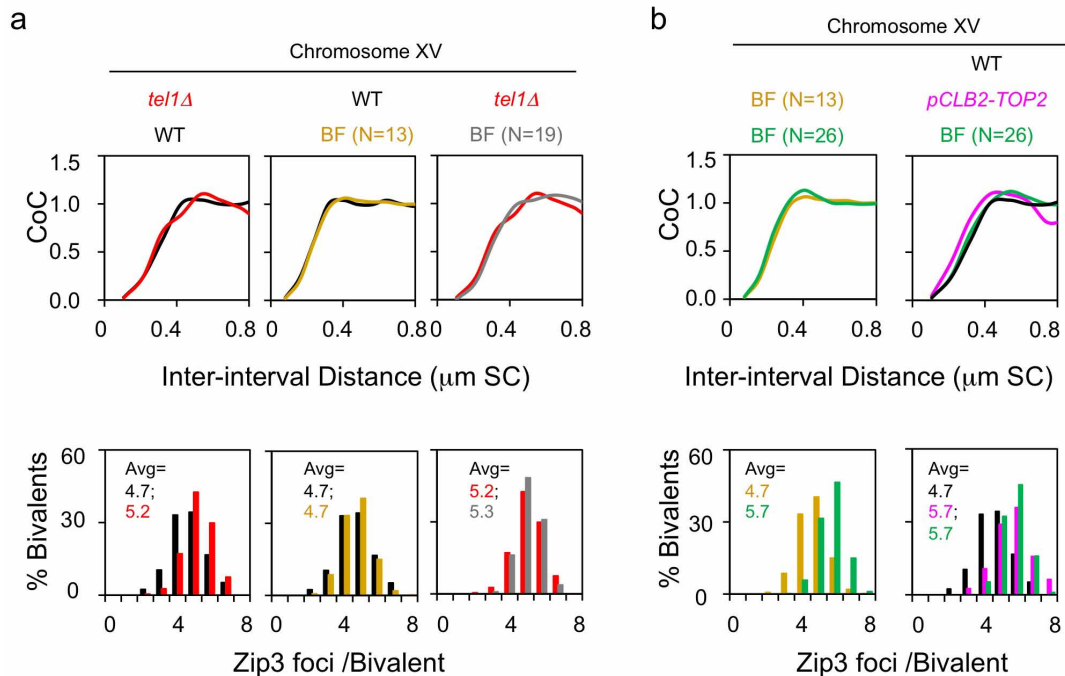


Extended Data Figure 5 | Increased level of SUMO–protein conjugates in *slx5Δ*. **a**, Western blots for whole protein extracts in WT and *slx5Δ* probed with anti-Smt3 antibody (Santa Cruz, sc-28649) and anti-Pgk1 antibody (Abcam ab113687) as a function of time after entry into meiosis ($t = 0$). Abundance of SUMO conjugates is increased in the mutant, especially in regions of high molecular mass. **b**, Quantification of the gel in **a**.



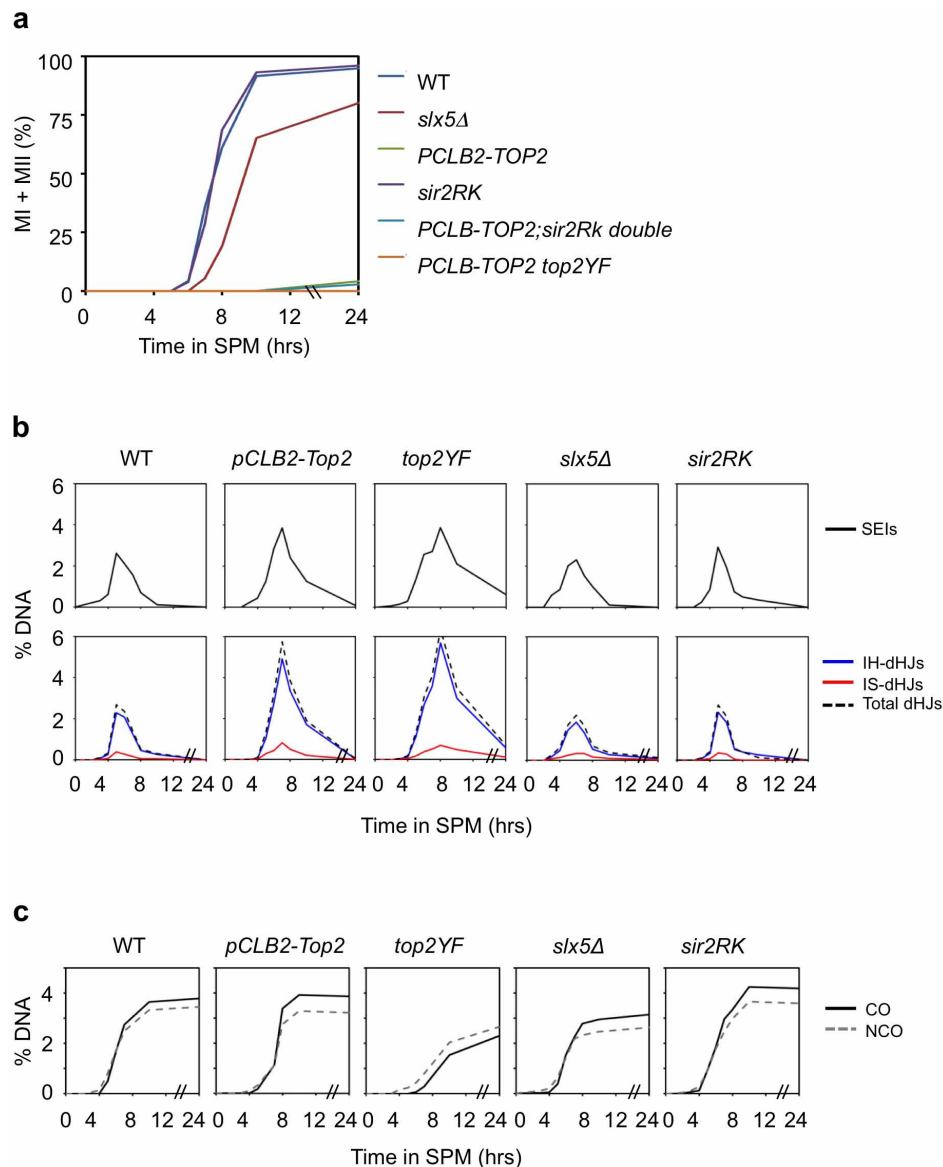
Extended Data Figure 6 | The role of Sir2 in crossover interference is specific to its interaction with Slx5. WT crossover interference is seen in diverse *sir2* non-null mutants affecting specific sub-functions (other than *sir2RK*; Fig. 3) and in mutants deleted for various interaction partners. *sir2-345*

is defective in histone deacetylase activity⁶³; *sir2ΔC500* lacks a Sir2 cohesion role⁶⁴. *sir3Δ*, *sir4Δ*, *esc2Δ* and *esc8Δ* eliminate Sir2 interaction partners involved in silencing^{43,65}; *hst1Δ* eliminates a Sir2 homologue⁶⁶.



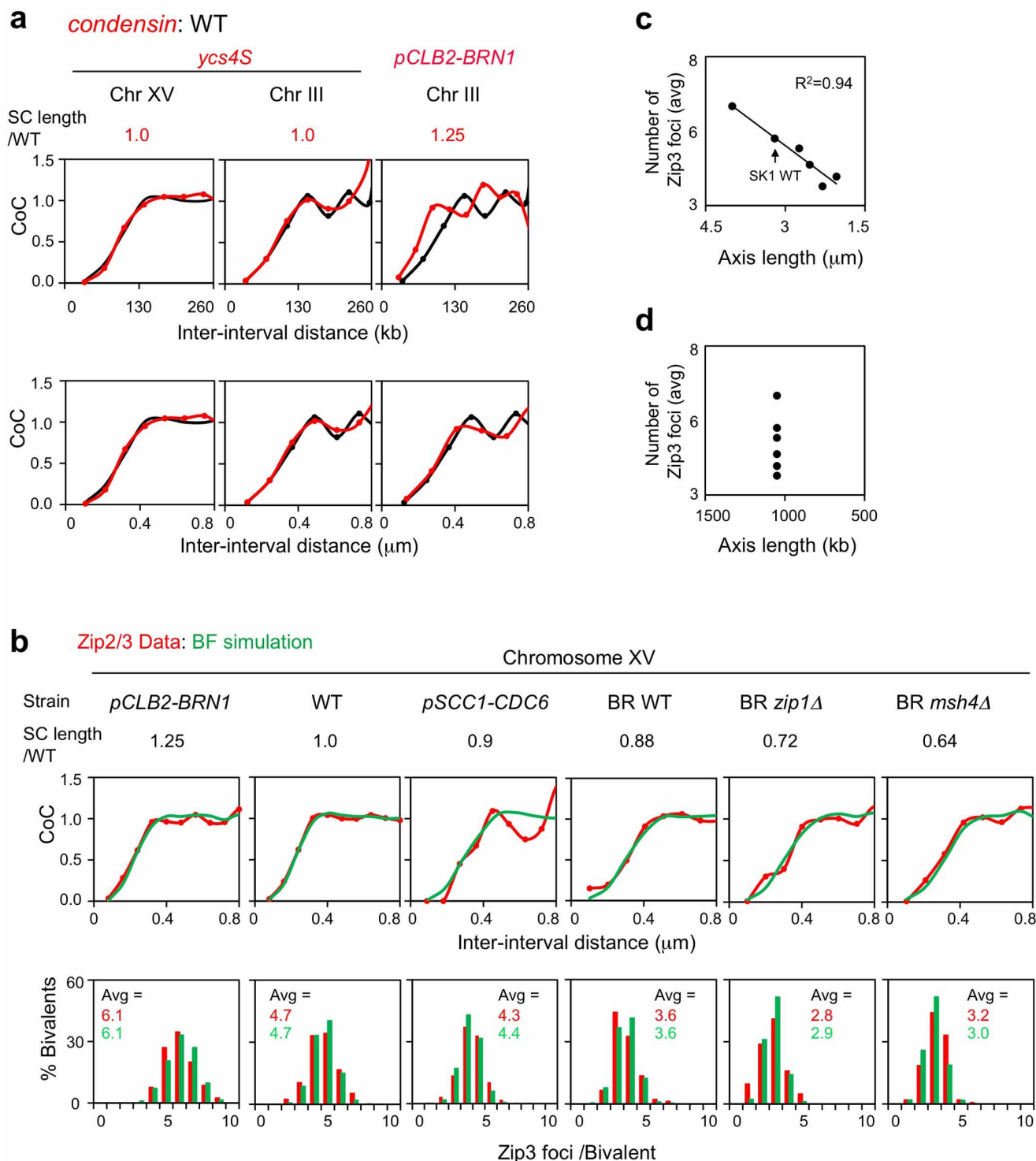
Extended Data Figure 7 | Mutant coefficient of coincidence and crossover number phenotypes cannot be explained by increased DSBs or by prolongation of the crossover-designation stage. Mutants in the described crossover interference pathway all confer coordinate changes in crossover interference, which is reduced, and the total number of crossovers, which is increased, by about 20% on chromosome XV. There are the expected consequences of a single defect in crossover interference, as illustrated by corresponding beam-film simulations, which quantitatively explain these results by a change in a single parameter, the interference length (L_{BF}) (Figs 2 and 3). This interference defect could comprise a defect in generation and spreading of the inhibitory signal and/or of the ability of unreacted precursors to respond to that signal (see text and Methods (section 'Beam-film simulations')). An increase in the number of crossovers can also occur as the result of either (1) prolongation of the crossover-designation period or (2) an increase in the number of DSBs⁸. Neither of these effects can explain the mutant phenotypes described in the text. (1) Crossover designation precedes synaptonemal complex formation and thus the pachytene stage¹⁴. Time-course analysis of representative mutant strains reveals that, in *sir2* mutants and in *top2SNM*, meiosis proceeds through pachytene and the two meiotic divisions normally (Extended Data Fig. 8a; ref. 14; data not shown). *slx5/8* mutants and *PCLB2-TOP2* mutants show no delay in progressing through prophase to pachytene (data not shown) but show a delay in meiosis I (*slx5*) or pachytene arrest (*PCLB2-TOP2*) (Extended Data Fig. 8a; data not shown). The *pCLB2-TOP2 top2YF* mutant does show a delay in achieving pachytene, as well as pachytene arrest, but exhibits the same crossover patterning phenotype as all other mutants, which show no pre-pachytene delay. Thus, prolonged crossover designation is not the basis for these phenotypes. (2) An increase in DSBs, without any change in crossover interference, does increase the number of crossovers; however, it has very little effect on crossover interference relationships (coefficient of coincidence curves) in budding yeast⁸. Correspondingly, two lines of evidence show that the mutant defects described here cannot be attributed to an increase in DSBs. **a**, A *tel1Δ* mutant exhibits

increased DSBs but no change in coefficient of coincidence relationships. *TEL1* encodes the yeast homologue of ATM. Absence of Tel1 confers a 50% increase in DSBs⁶² and a 10% increase in number of Zip3 foci (Supplementary Fig. 7 in ref. 8; reproduced in Extended Data Fig. 7a left, red colour). However, (1) there is no change in coefficient of coincidence relationships relative to WT (Extended Data Fig. 7a left), (2) the increase in crossovers is precisely that predicted on the basis of crossover homeostasis (ref. 8; text Fig. 2d, filled black circle at 19 DSBs/precursors per chromosome XV) and (3) beam-film simulation accurately describes the *tel1Δ* phenotype, relative to WT, by a change in a single parameter: the level of DSBs ($n = 19$, grey, versus 13, gold, in WT). The last point is documented in Extended Data Fig. 7a middle and right. The middle panel in Extended Data Fig. 7a shows the beam-film best-fit simulation for WT chromosome XV, where $n = 13$ (gold), compared with the experimental coefficient of coincidence curve (black; from Fig. 1); the right panel shows the beam-film best-fit simulation for *tel1Δ* chromosome XV, where $n = 19$ (grey) and all other parameters are the same as for WT, compared with the experimental coefficient of coincidence curve (black; from Fig. 1). **b**, Beam-film simulations predict no/little change in coefficient of coincidence with increasing DSBs for yeast chromosome XV (data not shown). More specifically, to explain the increased number of crossovers observed in the analysed mutants, for example *pCLB2-TOP2*, the value of N required for beam-film simulations of chromosome XV would be 26 (double the WT value of $N = 13$). If beam-film simulations are performed under the same parameter values used for WT except that $N = 26$ instead of $N = 13$, the predicted coefficient of coincidence curve is unchanged compared with that predicted for WT (left panel, compare gold for $N = 13$ with green for $N = 26$). Correspondingly, the coefficient of coincidence curve predicted for $N = 26$ (green) matches the WT coefficient of coincidence curve (black) and is unlike the coefficient of coincidence curve for the mutant (pink) (right panel). Additional evidence that DSB number is not altered in *pCLB2-TOP2* versus *TOP2* is presented in Extended Data Figs 4 and 8.



Extended Data Figure 8 | Progression of meiosis and of recombination in interference-defective mutants. Representative mutants were examined for progression of meiotic divisions and for recombination at the previously characterized *HIS4LEU2* locus⁶⁷ (strains in Extended Data Table 1). **a**, Meiotic divisions. The first meiotic division occurs normally in $sir2RK$ (defective in interaction with *Slx5*); it is delayed in $slx5\Delta$ and is completely absent in $PCLB2-TOP2$ and $PCLB-TOP2 top2YF$ due to arrest at pachytene²³ (L.Z., unpublished observations). **b**, **c**, DNA events. The *HIS4LEU2* locus probably provides a direct readout of DNA events independent of the effects of interference. *HIS4LEU2* does not exhibit crossover homeostasis²⁴, which implies that it is not sensitive to crossover interference⁸. This feature presumably reflects the fact that this locus is a very strong DSB hot spot. A DSB occurs at this site in virtually every nucleus with a concomitant reduction in DSBs (and thus crossover precursors) at other positions in its vicinity (N.K., unpublished observations). This locus may also undergo early crossover designation, thus also dominating crossover interference patterns per se. Importantly, Zip3 foci are used for diagnosis of crossover interference relationships⁸. Zip3 foci form as a specific consequence of programmed crossover designation; they do not mark the sites of non-interfering crossovers, which exhibit an entirely different pattern along the chromosomes⁸. Furthermore, formation of Zip3 foci is upstream of, and thus insensitive to, defects in later events, including (1) major perturbations in the kinetics of recombination or the fidelity with which initiated events (crossover-fated and/or non-crossover-fated) proceed to their assigned fates (see, for example, ref. 14) or (2) the potential occurrence of additional DSBs due to delayed synaptonemal complex formation (discussion in refs 8 and 56).

Thus, none of the recombination aberrancies detected by physical analysis of recombination in the analysed mutants (below) is relevant to their crossover interference phenotypes. Correspondingly, although all mutants give exactly the same crossover patterns (interference and crossover number) as defined by Zip3 foci, the mutants vary widely with respect to DNA recombination phenotypes. The results below can be summarized to say that (1) absence of *Slx5/8-Sir2* STUbL activity has little, or only subtle, effect(s) on recombination, whereas (2) absence of *TopoII* or *TopoII* catalytic activity confers delays and aberrancies. **b**, DSBs, SEIs and dHJs. Progression through recombination is very similar to WT in $sir2RK$ and $slx5\Delta$. Both $PCLB2-TOP2$ and $PCLB-TOP2 top2YF$ exhibit a phenotype corresponding to delayed progression beyond the point of crossover designation: DSBs appear on time; however, DSBs, single-end invasions (SEIs) and double Holliday junctions (dHJs) all accumulate to higher than normal levels at later than normal times, implying delayed progression of crossover-designated DSBs to SEIs, and of SEIs to dHJs, where SEIs and dHJs are both crossover-specific intermediates¹⁴. There is no significant alteration in homologue-versus-sister bias in any of the four mutants, with inter-homologue dHJs predominating over inter-sister dHJs similarly to WT in all cases. **c**, Inter-homologue crossover (CO) and non-crossover (NCO) products. Inter-homologue crossover and non-crossover levels are very similar to WT in $PCLB2-TOP2$ and show variations relative to WT in the other mutants. A differential deficit of crossovers versus non-crossovers in $PCLB2-TOP2 top2YF$ suggests a specific defect in crossover maturation in this mutant.



Extended Data Figure 9 | The metric of crossover interference is physical axis length (micrometres). **a**, This study considered two different condensin mutants, *ycs4S* and *pCLB2-BRN1*. Axis length is normal in *ycs4S* and longer than normal in *pCLB2-BRN1*. Analysis presented for chromosome XV in *pCLB2-BRN1* (Fig. 5) was also done on chromosome III in that mutant background (right column), confirming that coefficient of coincidence relationships are WT when the metric is physical chromosome length but not when the metric is genomic distance. We similarly analysed chromosomes III and XV in the *ycs4S* background (left and middle columns), confirming WT coefficient of coincidence relationships by both metrics. **b**, Zip3 focus analysis for chromosome XV in the indicated strains (red; from Fig. 5) and beam-film simulation analysis (green). Best-fit simulations could be obtained for all strains using the same parameter values as for WT meiosis, including interference distance ($L_{BF} \approx 0.3 \mu\text{m}$), except that the number of precursors (N) had to be

varied linearly with axis length. For the indicated strains, from left to right, $N = 17, 13, 12, 10, 9$ and 8 . This result implies direct interplay between physical chromosome length (micrometres of synaptonemal complex) and DSB probability, as discussed elsewhere. **c**, **d**, For the mutant cases described in **b**, experimentally observed average numbers of Zip3 foci vary linearly with axis length (**c**). In contrast, different numbers of Zip3 foci are observed for the different strains despite the fact that chromosome XV has the same genomic length in all cases (**d**). We also note that the best fit simulation for BR *zip1Δ* had to include a 10% decrease in the 'efficiency of maturation of crossover-designated interactions', which, in the present context, implies that in a *zip1Δ* background there is a 10% reduction in either (1) the stability of a Zip3 focus under cytological spreading conditions at the absence of synaptonemal complex or (2) the probability that a crossover designation will give a Zip3 focus.

Extended Data Table 1 | Strains used in this study

Strains	Genotype
NKY4146	<i>HMR::LacO-URA3^r, URA3::CYC1p-LacI-GFP^r, ZIP3-13myc::Hygromycin</i>
NKY4147	<i>URA3::CYC1p-LacI-GFP^r, scp1(Ch XV telomere)::LacO-LEU2^r, ZIP3-13myc::Hygromycin</i>
NKY4148	<i>leu2::LacI-GFP::Clonat^r, tel4::226xLacO::Kan^r, ZIP3-13myc::Hygromycin</i>
LZY1842	as NKY4146, except <i>pCLB2-TOP2::KanMX^r</i>
LZY1570	as NKY4147, except <i>pCLB2-TOP2::KanMX^r</i>
LZY1845	as NKY4148, except <i>pCLB2-TOP2::KanMX^r</i>
LZY2306	as NKY4147, except <i>top2-SNM::KanMX^r</i>
LZY2190	as NKY4147, except <i>pCLB2-TOP2::KanMX^r, top2(Y782F)::URA3</i>
LZY2237	as NKY4147, except <i>ubc9-GFP::KanMX^r</i>
LZY2207	as NKY4147, except <i>red1::kanMX6^r, LEU2::pYI-red1KR</i>
LZY2262	as NKY4147, except <i>pCLB2-TOP2::KanMX^r, tel1D::KanMX^r</i>
LZY2194	as NKY4147, except <i>pCLB2-TOP2::KanMX^r, spo11-HA3His6::KanMX4^r</i>
LZY2187	as NKY4147, except <i>pCLB2-TOP2::KanMX^r, spo11-HA3His6::KanMX4/spo11(D290A)-HA3His6::KanMX4</i>
LZY2266	as NKY4147, except <i>pCLB2-TOP2::KanMX^r, spo11-HA3His6::KanMX4/spo11-(Y135F)-HA3His6::KanMX</i>
LZY2054	as NKY4147, except <i>slx5D::natMX^r</i>
LZY2418	as NKY4148, except <i>slx5D::natMX^r</i>
LZY1983	as NKY4147, except as <i>slx8D::natMX^r</i>
LZY2325	as NKY4147, except <i>slx5D::nat1::slx5-sim(1-4)::KanMX^r</i>
LZY2319	as NKY4147, except <i>slx8-SS::natMX^r</i>
LZY1572	as NKY4147, except <i>sir2D::KanMX^r</i>
LZY1667	as NKY4146, except <i>sir2D::KanMX^r</i>
LZY2166	as NKY4148, except <i>sir2D::KanMX^r</i>
LZY2012	as NKY4147, except <i>sir2D::KanMX4::Sir2-R139K::natMX^r</i>
LZY1756	as NKY4147, except <i>sir2-345::natMX^r</i>
LZY1702	as NKY4147, except <i>sir2-DC500::KanMX/sir2-DC500::natNT2</i>
LZY1516	as NKY4147, except <i>sir3D::LEU2^r</i>
LZY1723	as NKY4147, except <i>sir4D::KanMX/sir4::natNT2</i>
LZY2146	as NKY4147, except <i>esc2D::KanMX^r</i>
LZY1718	as NKY4147, except <i>esc8D::KanMX^r</i>
LZY1451	as NKY4147, except <i>hst1D::KanMX^r</i>
LZY1201	as NKY4147, except <i>ndj1D::KanMX^r</i>
LZY1446	as NKY4147, except <i>hta1-S128A^r, hta2-S128A^r</i>
LZY1986	as NKY4147, except <i>pCLB2-NSE2::KanMX^r</i>
LZY932	as NKY4147, except <i>dot1D::KanMX^r</i>
LZY2006	as NKY4147, except <i>smc6-9::NAT^r</i>
LZY1163	as NKY4147, except <i>ndt80D::LEU2^r, REC8-3HA::URA3/+, pCLB2BRN1::KANMX4^r</i>
LZY1325	as NKY4146, except <i>ndt80D::LEU2^r, REC8-3HA::URA3/+, pCLB2BRN1::KANMX4^r</i>
LZY1261	as NKY4147, except <i>ndt80D::KanMX^r, REC8-3HA::URA3/+, ycs4S^r</i>
LZY1364	as NKY4146, except <i>ndt80D::KanMX^r, REC8-3HA::URA3/+, ycs4S^r</i>
LZY1471	as NKY4146, except <i>pch2D::KanMX^r</i>
LZY1488	as NKY4148, except <i>pch2D::KanMX^r</i>
LZY1472	as NKY4147, except <i>pch2D::KanMX^r</i>
LZY773	as NKY4147, except <i>cdc6::kanMX6::PSCC1:3-HA-CDC6^r, ndt80::LEU2^r</i>
LZY1317	as NKY4147, except <i>mlh1D::KanMX^r</i>
LZY1386	as NKY4147, except <i>mlh3D::KanMX^r</i>
LZY1318	as NKY4147, except <i>mms4D::KanMX^r</i>
LZY1504	as NKY4147, except <i>msh2::LEU2^r</i>
LZY2018	as NKY4147, except <i>sir2D::KanMX4::Sir2-R139K::nat^r, pCLB2-TOP2::KanMX^r</i>
LZY2080	as NKY4147, except <i>sir2D::KanMX4::Sir2-R139K::nat^r, slx5D::natMX^r</i>
LZY2313	as NKY4147, except <i>slx5D::natMX^r, red1::kanMX6^r, LEU2::pYI-red1KR,</i>
LZY2430	as NKY4147, except <i>slx5D::natMX^r, top2-SNM::KanMX^r</i>
LZY2341	as NKY4147, except <i>top2-SNM::KanMX, red1::KanMX, LEU2-red1KR</i>
LZY446	<i>ho::hisG leu2 ura3 nuc1::hygroB HIS4::LEU2-(BamHI+ori), MAT alpha</i>
LZY447	<i>ho::hisG leu2 ura3 nuc1::hygroB his4-x::LEU2-(NgoMIV+ori)--URA3, MAT a</i>
LZY1614	as LZY446, except <i>pCLB2-TOP2::KanMX</i>
LZY1617	as LZY447, except <i>pCLB2-TOP2::KanMX</i>
LZY2413	as LZY446, except <i>pCLB2-TOP2::KanMX, URA3::top2(Y782F)</i>
LZY2414	as LZY447, except <i>pCLB2-TOP2::KanMX, URA3::top2(Y782F)</i>
LZY2261	as LZY446, except <i>slx5D::natMX</i>
LZY2255	as LZY447, except <i>slx5D::natMX</i>
LZY2198	as LZY447, except <i>sir2D::KanMX4::Sir2-R139K::nat</i>
LZY2199	as LZY447, except <i>sir2D::KanMX4::Sir2-R139K::nat</i>

All strains are isogenic derivatives of SK1 with *ho::hisG, leu2* and *ura3*.

Structure of class C GPCR metabotropic glutamate receptor 5 transmembrane domain

Andrew S. Doré^{1*}, Krzysztof Okrasa^{1*}, Jayesh C. Patel^{1*}, Maria Serrano-Vega^{1*}, Kirstie Bennett¹, Robert M. Cooke¹, James C. Errey¹, Ali Jazayeri¹, Samir Khan¹, Ben Tehan¹, Malcolm Weir¹, Giselle R. Wiggin¹ & Fiona H. Marshall¹

Metabotropic glutamate receptors are class C G-protein-coupled receptors which respond to the neurotransmitter glutamate. Structural studies have been restricted to the amino-terminal extracellular domain, providing little understanding of the membrane-spanning signal transduction domain. Metabotropic glutamate receptor 5 is of considerable interest as a drug target in the treatment of fragile X syndrome, autism, depression, anxiety, addiction and movement disorders. Here we report the crystal structure of the transmembrane domain of the human receptor in complex with the negative allosteric modulator, mavoglurant. The structure provides detailed insight into the architecture of the transmembrane domain of class C receptors including the precise location of the allosteric binding site within the transmembrane domain and key micro-switches which regulate receptor signalling. This structure also provides a model for all class C G-protein-coupled receptors and may aid in the design of new small-molecule drugs for the treatment of brain disorders.

Glutamate is the major excitatory neurotransmitter in the central nervous system. It mediates its activity through ionotropic channels and eight metabotropic G-protein-coupled receptors (GPCRs) which are expressed in neuronal and glial cells. The metabotropic glutamate (mGlu) receptors can be divided into three groups on the basis of their sequence similarity, pharmacology and transduction mechanisms. mGlu₅ and mGlu₁ are class I mGlu receptors which are primarily located postsynaptically and couple to the G_{q/11} pathway¹. mGlu₅ is abundant throughout the cortex, hippocampus, striatum, caudate nucleus and nucleus accumbens², areas involved in emotion, motivation and cognition. mGlu₅ is a promising therapeutic target and negative allosteric modulators (NAMs) which reduce mGlu₅ receptor activation are undergoing clinical trials for the treatment of fragile X syndrome, depression, anxiety,

migraine, and dyskinesias^{3,4}, whereas positive allosteric modulators (PAMs) which increase receptor activation may be beneficial in schizophrenia and cognitive disorders⁵.

GPCRs can be divided into four major classes (A, B, C and F) on the basis of sequence similarity⁶. Structures of over 20 class A receptors^{7,8}, one class F⁹, and two class B receptors^{10,11} have been solved to date. Class C GPCRs include mGlu receptors, γ -aminobutyric acid B-type receptors (GABA_B), calcium-sensing receptors (CaS), taste receptors (TAS) and several orphan receptors^{1,12}. mGlu receptors have an unusual structure comprising a large extracellular domain consisting of the 'venus fly trap' (VFT), which binds glutamate and a cysteine-rich domain (CRD), linked to the seven-transmembrane domain (TMD). A characteristic of class C GPCRs is that they exist as dimers, which in the case of the mGlu

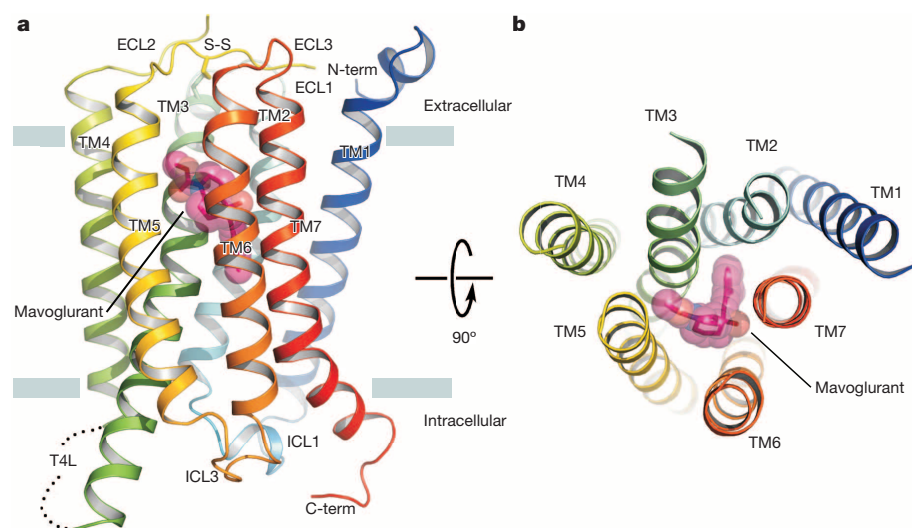


Figure 1 | Schematic and ribbon representation of the mGlu₅ structure. **a**, **b**, Ribbon representation of mGlu₅ in rainbow colouration (N terminus, blue; C terminus, red) viewed parallel to the membrane and from the extracellular space, respectively. Mavoglurant is represented in translucent stick representation. Carbon, nitrogen and oxygen atoms are coloured magenta, blue and red, respectively. The position of the T4L-insertion to ICL2 is indicated.

¹Heptares Therapeutics Ltd, BioPark, Broadwater Road, Welwyn Garden City, Hertfordshire AL7 3AX, UK.

*These authors contributed equally to this work.

family is mediated by interactions between the VFT and TMDs. Crystal structures of several VFT domains have been solved^{13,14}, however, structural information on the TMD has until recently remained elusive. This domain is of particular importance as drug discovery efforts have focused on the identification of allosteric modulators which bind within this region and allow greater subtype selectivity^{12,15,16}. To advance our understanding of the mode of action of class C GPCRs and to enable structure-based drug design of allosteric modulators, we have determined the crystal structure of the TMD of the human mGlu₅ receptor in complex with the NAM mavoglurant (AFQ056)⁴, which is currently in phase III clinical trials for fragile X syndrome.

Structure determination

To determine the TMD structure of mGlu₅, a thermostabilized receptor (StaR) was generated as previously described^{17–19}. The receptor was thermostabilized in the presence of the allosteric radioligand 2-methyl-6-((3-methoxyphenyl)ethynyl)-pyridine ([³H]-M-MPEP)²⁰ (Extended Data Fig. 1a); and contains six mutations, none of which are located in the allosteric binding site (Extended Data Fig. 1b).

To facilitate crystallization, the flexible domains were removed from the N terminus (residues 2–568), and C terminus (residues 837–1153, including 2 residues from the predicted helix 8), and T4-lysozyme (T4L) inserted into intracellular loop (ICL) 2 between Lys 678 and Lys 679 (Fig. 1a). The affinity of [³H]-M-MPEP for this construct (mGlu₅StaR (569–836)-T4L) does not differ from the wild-type receptor ($K_d = 0.86 \pm 0.04$ nM and 1.05 ± 0.15 nM, respectively; Extended Data Table 1) and there is no difference in the affinities of a range of allosteric modulators for the wild-type and StaR (Extended Data Table 1 and Extended Data Fig. 2). Modifications of the protein required for crystallization preclude G protein coupling. In the absence of the glutamate-binding domain, mavoglurant cannot act as an allosteric modulator.

The structure of mGlu₅ bound to methyl (3aR,4S,7aR)-4-hydroxy-4-[(3-methylphenyl)ethynyl]octahydro-1*H*-indole-1-carboxylate (mavoglurant) was determined to 2.6 Å through merging diffraction data from 5 crystals grown in lipidic cubic phase (LCP). The structure was solved by molecular replacement with one copy of the receptor in the asymmetric unit (Extended Data Figs 3 and 4). Details of data collection and refinement are in Extended Data Table 2.

Overall architecture of mGlu₅

The TMD of mGlu₅ is comprised of seven-transmembrane helices (TM1–7) (Fig. 1a, 1b). Residues are assigned numbers (in superscript) based on a modification of the Ballesteros-Weinstein system²¹ previously suggested¹ for class C (Extended Data Table 3). Continuous density is observed for the intracellular loops ICL1 (forming a short α -helix), ICL3 and for the extracellular loops (ECL) 1 and 3, which lack secondary structure. Cys 644^{3,29} at the N-terminal end of TM3 forms a conserved disulphide bond with Cys 733 in ECL2 (Fig. 1a). ECL2 spans the top of the receptor interacting with the N terminus of TM1 (Tyr 730 to Asp 577), the C terminus of TM2 (Val 729 to Ala 637^{2,58} and Leu 731 to Leu 635^{2,56}), ECL1 and TM3 (Leu 731 to Gln 647^{3,32}). The disulphide bond between the top of TM3 and ECL2 is critical in anchoring ECL2, however, the interaction with Gln 647^{3,32} also appears conserved with glutamine or arginine at the equivalent position in 92% of class C receptors (percentage conservation values used refer to vertebrate class C receptors as shown in Extended Data Table 3).

The configuration of the helical bundle and the position of ECL2 combine to severely restrict the entrance to the allosteric pocket (Fig. 2). This is consistent with native ligands binding to the extracellular domain, and as such there is no requirement for a wide entrance to the binding pocket in the transmembrane domain. At the C terminus of the receptor, an outward kink at the end of TM7 is propagated by Pro 820^{7,50} and Lys 821^{7,51}; these two residues are highly conserved (> 90%) across class C receptors, suggesting that the helix kink may be a common structural feature in class C.

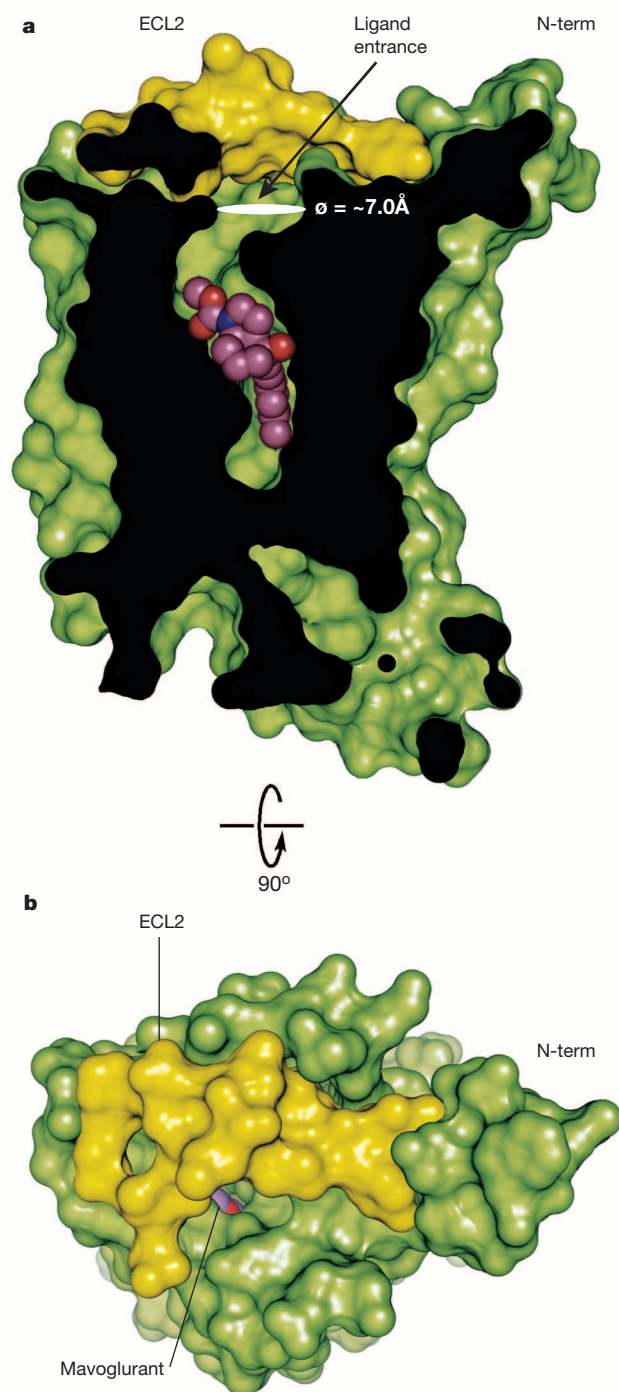


Figure 2 | The narrow entrance to the allosteric pocket in mGlu₅. Surface representation of the mGlu₅ TMD from two orientations separated by 90°. **a**, View parallel to the membrane plane. **b**, View from the extracellular space. The surface of ECL2 is coloured yellow, and the rest of the receptor green. Mavoglurant is represented in a space-filling representation, with carbon, nitrogen and oxygen atoms coloured magenta, blue and red, respectively. The allosteric pocket exhibits a narrow opening (ellipsoid denotes theoretical entrance dimensions at widest point in **a**) as a result of TM helix conformations and a network of interactions anchoring ECL2 across the top of the receptor.

Comparison with class A and B GPCRs

The superfamily of GPCRs is presumed to have evolved from a common ancestral receptor as multiple sequence motifs are shared across two or more subfamilies²². With the mGlu₅ structure in hand, we could explore a structural alignment between class A, B and C subfamilies.

Comparisons with mGlu₅ have been performed using rhodopsin (PDB ID: 1F88), which shows the closest alignment from class A²³, and the CRF₁R structure (PDB ID: 4K5Y)¹⁰ from class B, with all structures in the inactive state.

Global (all atom) superposition of the mGlu₅ structure with rhodopsin and CRF₁R, and local superposition of TM helices (Extended Data Table 4) reveal that the consensus across TM positions appears best across the intracellular halves of the receptors (Fig. 3a–f), consistent with the structural constraints of G protein coupling, and with the highest levels of structural diversity observed across the extracellular portions. TM1 is structurally well aligned across all three receptor classes, differing most at the N termini. In class A and B receptors, similar sets of interactions between TM1 and TM7 influence the trajectory of the extracellular portion of TM7. In mGlu₅ this is closest to the TM7 of rhodopsin, yet shifted by 5 Å towards the centre of the helical bundle, and towards TM2 contributing to the narrow entrance to the allosteric pocket (Fig. 3c).

The extracellular trajectory of TM2 in mGlu₅ more closely resembles that of CRF₁R (Fig. 3f), appearing straighter and without the bend towards TM1 observed in several class A structures. However, the extracellular half of TM2 in mGlu₅ is shifted closer to the central axis of the receptor in comparison to CRF₁R by ~2.5 Å due to interaction with the

highly conserved Tyr 629^{2,50} (89% tyrosine or phenylalanine) packing against Gly 590^{1,50} (90% conserved), and towards TM7. The centre of TM2 in mGlu₅ is closer to the central axis, which constricts the extracellular side and moves the intracellular end of TM2 further out from the helical bundle.

The most striking difference between mGlu₅, rhodopsin and CRF₁R is the position of TM5. Though the extracellular portion of TM5 in mGlu₅ grossly follows that of CRF₁R, the entire helix is positioned further inwards in comparison to both rhodopsin and CRF₁R (Fig. 3b, c, e, f) by approximately 6 Å which further contributes to the narrow entrance to the allosteric cavity.

Interactions in conserved motifs

The TMD of mGlu₅ is considered to function in a similar way to class A receptors²⁴. The isolated monomeric TMD of mGlu₂ couples to G proteins in response to a PAM²⁵. GPCRs exhibit a set of conserved microdomains that are important in signal transduction^{26,27}. One highly conserved motif is the E(D)RY sequence in TM3, termed the 'ionic lock', which bridges TM3 and TM6. In rhodopsin, a salt bridge is formed between Arg 135^{3,50} and Glu 247^{6,30} and is considered a hallmark of the inactive state. Similarly, in the mGlu₅ structure Lys 665^{3,50} in TM3 forms a salt bridge with Glu770^{6,35} (2.7 Å distance) in TM6, and interacts with

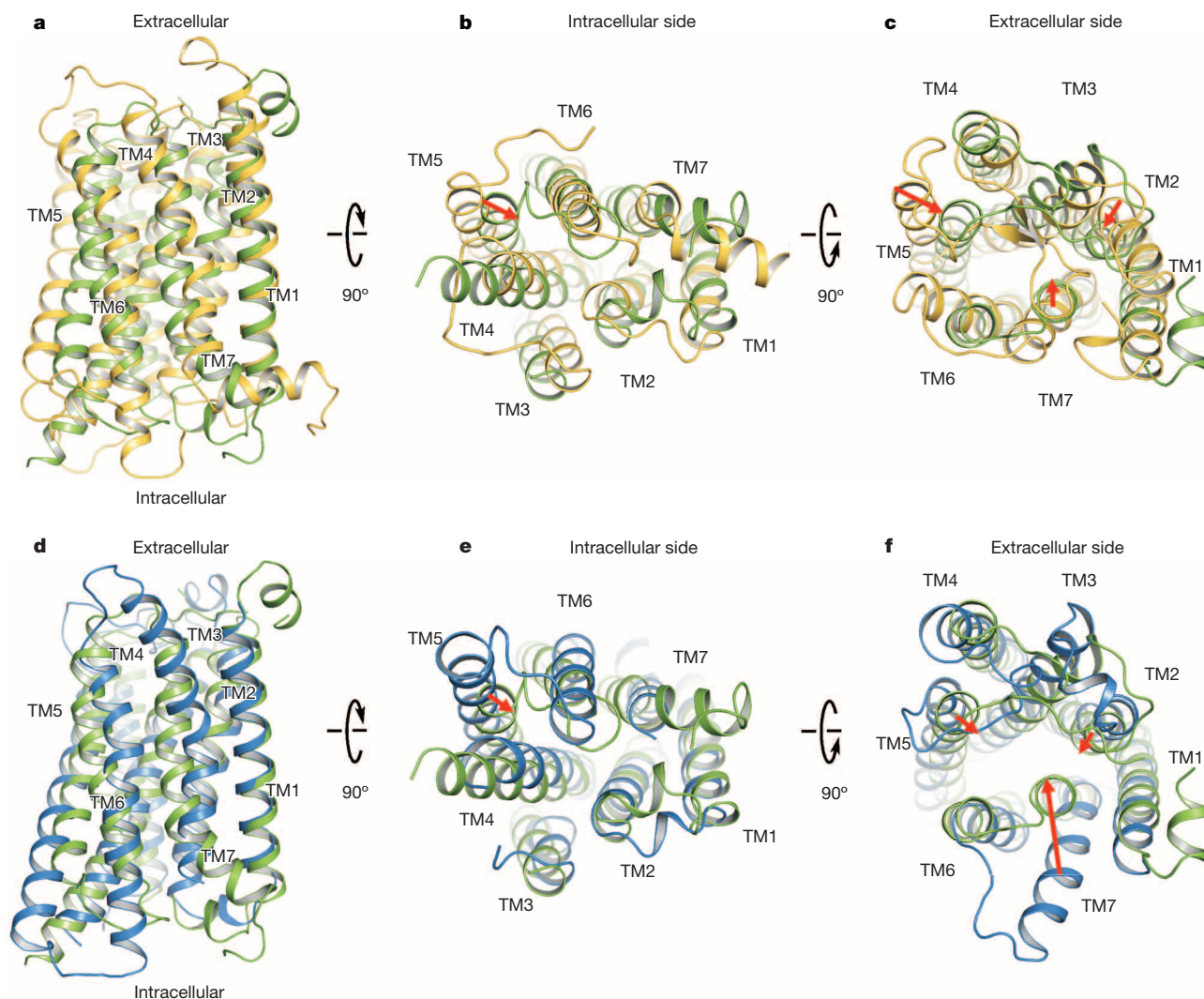


Figure 3 | Superposition and comparison of mGlu₅ with rhodopsin and CRF₁R. The superimposed structures of mGlu₅, CRF₁R (PDB ID: 4K5Y) and rhodopsin (PDB ID: 1F88) are represented as green, blue and yellow ribbons, respectively. **a–c**, mGlu₅ superposed with rhodopsin is viewed parallel to the membrane plane (**a**), from the intracellular side (**b**) and extracellular side (**c**).

d–f, mGlu₅ superposed with CRF₁R is viewed parallel to the membrane plane (**d**), from the intracellular side (**e**) or from the extracellular side (**f**). For both superpositions, rotations indicated for panels **b**, **e** and **c**, **f** are relative to **a**, **d**, respectively. Red arrows denote large differences in transmembrane positions between class C, class A and class B receptors.

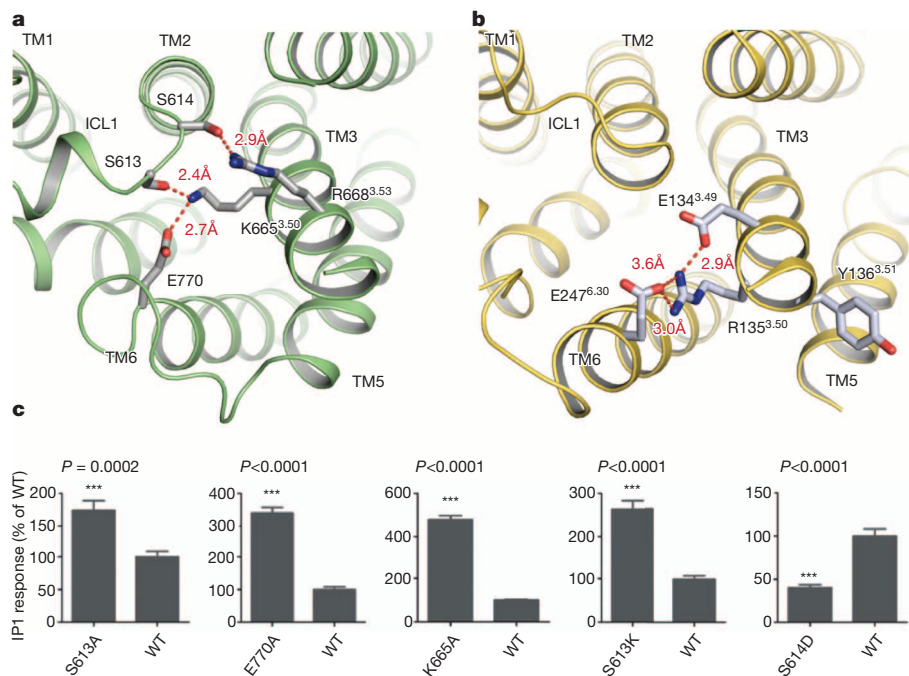


Figure 4 | The 'ionic-lock' motifs in mGlu₅ and rhodopsin. **a, b**, mGlu₅ shown as green ribbons (**a**) and rhodopsin as yellow ribbons (**b**). Interacting residues are shown as sticks with carbon, nitrogen and oxygen atoms coloured grey, blue and red, respectively. Specific interactions depicted as dashed red lines (distances labelled). **c**, Constitutive activity of mGlu₅ mutants compared to wild type (WT). IP1 response is expressed as percentage of wild type, after normalization to number of receptors per cell for each construct (see Extended Data Fig. 5). Error bars indicate standard error of mean, *P* values derived from an unpaired two-tail *t*-test. Data are representative of three independent experiments.

Ser 613 (2.4 Å distance) in ICL1 (Fig. 4a, b). Additionally, the less well conserved Arg 668^{3.53} is within hydrogen bonding distance of Ser 614 in ICL1 (3.1 Å distance), thus making a secondary lock further tethering TM6 to TM3 via ICL1 (Fig. 4a). Two of the stabilizing mutations flank Arg 668^{3.53} (N667Y and I669A). However, Tyr 667 faces outwards and is unlikely to influence this network of interactions.

Consistent with a role in maintaining an inactive conformation, disruption of the ionic lock through mutation of Ser 613, Glu 770^{6.35} or Lys 665^{3.50} to alanine significantly increases constitutive activity compared to wild-type (Fig. 4c). Mutating Ser 613 to lysine also results in higher levels of constitutive activity caused by charge repulsion with Lys 665^{3.50}. Importantly, substituting Ser 614 with aspartic acid to create a stronger ionic interaction with Arg 668^{3.53} than the naturally occurring hydrogen bond, results in boosting this secondary lock and a significant decrease in constitutive activity (Fig. 4c). Taken together, these results provide functional evidence for the role of this network in modulating receptor activity.

Lys 665^{3.50}, Glu 770^{6.35} and Ser 613 are highly conserved (>77%) across class C GPCRs including mGlu, GABA_B, CaS and TAS1 receptors. In both GABA_B subunits, the glutamate is substituted with aspartate. In the TAS1 taste receptors these residues vary in that in TAS1R3 the glutamate and arginine in TM6 are swapped in position and the lysine is replaced with asparagine. In TAS1R1/TAS1R2, Lys 665^{3.50} is replaced with arginine and glutamine is found at the equivalent position to Arg 668^{3.53}. Mutations in one or more of these residues have been found to alter the signalling of GABA_{B2}, the signalling subunit of the GABA_B receptor²⁸ and the CaS receptor²⁹. This region is also important in human disease as the mutation E781K^{6.35} in mGlu₆ leads to congenital night blindness and altered G protein coupling³⁰. Collectively, these data highlight the functional importance of this network within class C receptors.

Using the mGlu₅ structure, it is possible to rationalize the role of other residues highly conserved across class C receptors. Leu 622^{2.43} (92% conserved) in TM2 points inwards towards a conserved aromatic cage formed by residues in TM1, TM2 and TM7. Leu 750^{5.50} (100% conserved) and Tyr 746^{5.46} (83% conserved) in TM5 appear structurally important in stabilizing TM3, with Tyr 746^{5.46} additionally bridging TM5 to TM4. The hydroxyl group of Tyr 746^{5.46} interacts with the backbone carbonyl of Leu 700^{4.39} in TM4, breaking the helical hydrogen bonding pattern and influencing the trajectory of the extracellular portion of TM4.

The conserved NP^{7.50}xxY(x)₅F motif, found at the junction of TM7 with helix 8 in class A receptors, undergoes substantial rearrangements

during activation^{31,32}. The conserved tyrosine (Tyr 306^{7.53} in rhodopsin) faces inwards, moving to fill the space created by the outward movement of TM6 upon activation. In mGlu₅ the conserved motif F/Y/HxPKxY found at the intracellular end of TM7 appears to play an analogous role but in a different way (Fig. 5a, b). First, Tyr 823^{7.53} faces out towards the membrane, C-terminal from the kink at the intracellular end of TM7, and ~8.3 Å away from Tyr 306^{7.53} in rhodopsin (measured between Cα positions). It is therefore unlikely that Tyr 823^{7.53} in mGlu₅ performs a similar role to that of the conserved tyrosine in the NPxxY(x)F motif during class A activation. Instead, the highly conserved Lys 821^{7.51} and/or

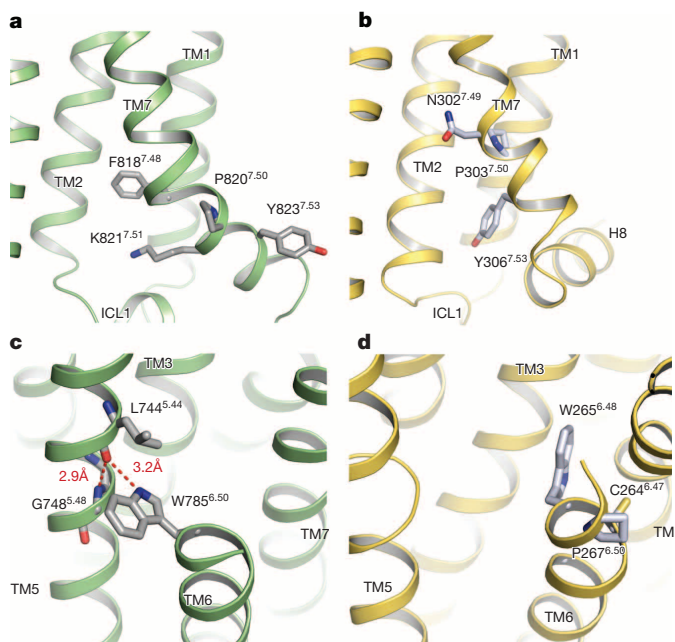


Figure 5 | Molecular details of further conserved sequence motifs in mGlu₅ and comparison to rhodopsin. **a–d**, Residues in the FxPKxY and NPxxY motifs (**a, b**) and residues in the FxxCWxP motif (**c, d**), in mGlu₅ and rhodopsin, respectively. mGlu₅ and rhodopsin colour schemes as in Fig. 4. Interacting residues are shown as sticks with carbon, nitrogen and oxygen atoms coloured grey, blue and red, respectively. Specific interactions are depicted as dashed red lines with distances labelled.

Phe 818^{7,48} in mGlu₅ superpose across the position of Tyr 306^{7,53} in rhodopsin with both residues capable of stabilizing the inter-helical space created by the outward movement of TM6.

Finally, the highly conserved FxxCWxP^{6,50} motif in TM6 of rhodopsin contains the tryptophan at position 6.48 known as the toggle switch, which alters rotamer states upon activation. A kink induced in TM6 by Pro 267^{6,50} in rhodopsin, along with rotamer changes of the tryptophan, form the basis for the outward movement of the intracellular end of TM6 upon activation. Class C receptors also have a highly conserved (83%) tryptophan in the equivalent position in TM6. However, in the mGlu₅ structure, TM6 has moved laterally away from TM7 compared to rhodopsin by 3.3 Å. A potential bifurcated hydrogen bond from the main-chain carbonyl of Leu 744^{5,44} to Gly 748^{5,48} and the indole ring of Trp 785^{6,50} then bridges TM6 to TM5, with the observed rotation of the tryptophan side chain proving critical for the observed binding position of mavoglurant (Fig. 5c, d).

Allosteric modulator binding site

Strong density for mavoglurant (Fig. 6b) is found in a pocket ~8 Å from the receptor surface and spanning ligand positions observed previously for class A and B (Fig. 6c). The ligand is oriented with an approximate 30° tilt (relative to the axis of the TM bundle) in a pocket defined by residues from TM2, TM3, TM5, TM6 and TM7 (Fig. 6a).

The 3-methylphenyl ring of mavoglurant sits in a pocket between Ala 810^{7,40} and Pro 655^{3,40}, bordered on one side by Ile 625^{2,46}, Gly 628^{2,49}, Ser 654^{3,39}, Ser 658^{3,43} and on the other by Tyr 659^{3,44}. The alkyne linker traverses a narrow channel between Tyr 659^{3,44}, Ser 809^{7,39}, Val 806^{7,36} and Pro 655^{3,40}, with the hydroxyl substituent at the other end sitting almost equidistant between, and within hydrogen bonding distance of, the hydroxyl oxygens from Ser 809^{7,39} and Ser 805^{7,35} (Fig. 6a, b). The shape and interactions in this region of the binding pocket partly explains the predominance of an alkyne linker in mGlu₅ allosteric modulators. The saturated bicyclic ring system sits within a mainly hydrophobic pocket defined by Val 806^{7,36}, Met 802^{7,32}, Phe 788^{6,53}, Trp 785^{6,50}, Leu 744^{5,44}, Ile 651^{3,36}, Pro 655^{3,40} and Asn 747^{5,47}. The Asn 747^{5,47} δ-nitrogen hydrogen bonds to the carbonyl oxygens of Gly 652^{3,37} and the carbamate tail of mavoglurant, with the methoxy terminal portion of the carbamate sitting in a cavity defined by Leu 744^{5,44}, Pro 743^{5,43} and Ile 651^{3,36}. The importance of many of these residues in allosteric ligand binding to mGlu₅, has been suggested previously in extensive mutagenesis studies^{33–36}.

In the mGlu₅ structure the 3-methyl substituent of mavoglurant is within 4 Å of a network of hydrogen bonds involving the side chains of Tyr 659^{3,44}, Thr 781^{6,46}, the main-chain carbonyl of Ser 809^{7,39} and a water molecule at the bottom of the allosteric pocket (Fig. 6b). Varying the 3-methyl substituent on the phenyl ring in related compounds from methoxy to chloro to fluoro switches the ligand from a NAM to a neutral binder to a PAM, respectively³⁷. Furthermore, mutation of Thr 781^{6,46} (Thr 780_{rat}) and Ser 809^{7,39} (Ser 808_{rat}) to alanine also switches the pharmacology of alkyne type PAMs³⁵. Although the side chain of Thr 781^{6,46} is directly involved in the hydrogen bond network described above, the side chain of Ser 809^{7,39} makes a hydrogen bond to the carbonyl of Ser 805^{7,35} in addition to mavoglurant. This induces a small bend in TM7 (ref. 38) and orients the main-chain carbonyl of Ser 809^{7,39} to engage with the water (Fig. 6b). The observed changes in ligand pharmacology resulting from perturbation of the network around the water, point to a potential activation switch holding TM7, TM3 and TM6 together at the base of the pocket, and a structural basis for modulation of the conformational spectra of the receptor.

While this manuscript was under review, the structure of the transmembrane domain of the related class C receptor mGlu₁ in complex with a different class of NAM 4-fluoro-N-(4-(6-(isopropylamino)pyrimidin-4-yl)thiazol-2-yl)-N-methylbenzamide (FITM) was published³⁹. The two structures are in close agreement, with a number of the molecular features described here conserved across subtypes. However, in comparison to the position of mavoglurant in mGlu₅, FITM is found higher in the mGlu₁ allosteric site, a position more analogous to that of compounds bound in the orthosteric site of class A receptors, highlighting the potential for different allosteric modulator binding modes across this class of receptors.

Conclusions

The structure of the TMD of the mGlu₅ receptor provides a detailed view of the core domain of a class C GPCR, complementing the extracellular domain structures. Alongside structures for class A, B and F receptors, this completes initial structural coverage of the classes of GPCRs, enabling conserved and divergent structural and mechanistic features to be identified. The mGlu₅ receptor structure also increases our understanding of the mechanism of action of allosteric modulators for metabotropic receptors and will enable the design of both negative and positive allosteric modulators by providing a template for homology

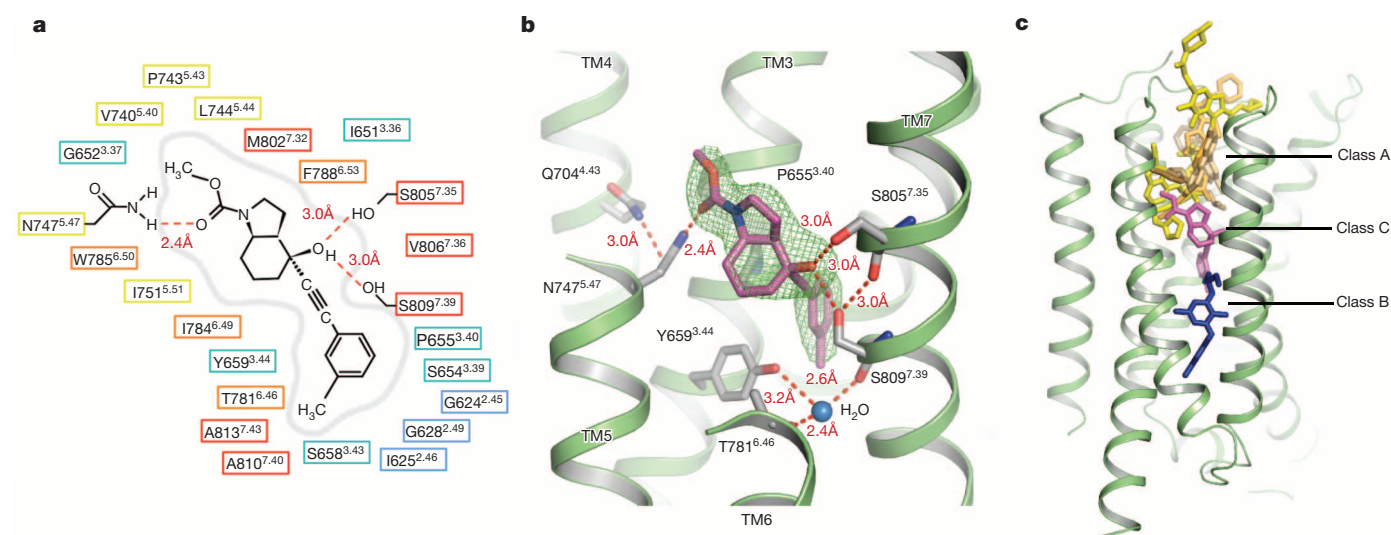


Figure 6 | The mGlu₅ allosteric modulator binding site. **a**, Diagram of ligand interactions in the binding pocket. The colour scheme of the boxes follows the rainbow colouration as in Fig. 1b. Hydrogen bonds depicted as dashed red lines with distances between heavy atoms in angstroms. **b**, View from the membrane of specific interactions in the binding pocket. Extracellular portions

of TM5/TM6 removed for clarity. $F_o - F_c$ OMIT density contoured at 2.0 σ calculated before mavoglurant inclusion in the model, mavoglurant shown as sticks as in Fig. 1a. **c**, The location of mavoglurant (magenta) in comparison to an antagonist (blue) bound to CRF1R, and selection of ligands (yellow) bound to class A receptors.

modelling of related receptors. Drugs directed at this class of receptors show great promise in the treatment of severe neuropsychiatric disorders.

METHODS SUMMARY

Conformationally thermostabilized human mGlu₅ lacking the N-terminal extracellular domain and with T4-lysozyme fused into intracellular loop 2 was expressed in baculovirus-infected insect cells and purified in *n*-dodecyl- β -D-maltopyranoside (DDM) using a histidine-tag. Crystals were grown in LCP at 20.0 °C using a mono-olein/cholesterol mixture. Diffraction data from 5 crystals, collected at Diamond Light Source beamline I24, was used to solve the structure by molecular replacement using the structures of T4-lysozyme and an ensemble of 8 GPCR TMD structures as initial search models. The structure was refined to 2.6 Å with good statistics ($R_{\text{work}}/R_{\text{free}} = 23.9/27.4\%$). For further details of experimental procedures see Methods.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 January; accepted 22 April 2014.

Published online 6 July 2014.

- Pin, J. P., Galvez, T. & Prézeau, L. Evolution, structure, and activation mechanism of family 3/C G-protein-coupled receptors. *Pharmacol. Ther.* **98**, 325–354 (2003).
- Shigemoto, R. *et al.* Immunohistochemical localization of a metabotropic glutamate receptor, mGluR5, in the rat brain. *Neurosci. Lett.* **163**, 53–57 (1993).
- Li, G., Jørgensen, M. & Campbell, B. M. Metabotropic glutamate receptor 5-negative allosteric modulators for the treatment of psychiatric and neurological disorders (2009–July 2013). *Pharm. Pat. Anal.* **2**, 767–802 (2013).
- Levenga, J. *et al.* AFQ056, a new mGluR5 antagonist for treatment of fragile X syndrome. *Neurobiol. Dis.* **42**, 311–317 (2011).
- Stauffer, S. R. Progress toward positive allosteric modulators of the metabotropic glutamate receptor subtype 5 (mGlu5). *ACS Chem. Neurosci.* **2**, 450–470 (2011).
- Venkatakrishnan, A. J. *et al.* Molecular signatures of G-protein-coupled receptors. *Nature* **494**, 185–194 (2013).
- Rosenbaum, D. M., Rasmussen, S. G. F. & Kobilka, B. K. The structure and function of G-protein-coupled receptors. *Nature* **459**, 356–363 (2009).
- Katritch, V., Cherezov, V. & Stevens, R. C. Diversity and modularity of G protein-coupled receptor structures. *Trends Pharmacol. Sci.* **33**, 17–27 (2012).
- Wang, C. *et al.* Structure of the human smoothened receptor bound to an antitumour agent. *Nature* **497**, 338–343 (2013).
- Hollenstein, K. *et al.* Structure of class B GPCR corticotropin-releasing factor receptor 1. *Nature* **499**, 438–443 (2013).
- Siu, F. Y. *et al.* Structure of the human glucagon class B G-protein-coupled receptor. *Nature* **499**, 444–449 (2013).
- Bräuner-Osborne, H., Wellendorf, P. & Jensen, A. A. Structure, pharmacology and therapeutic prospects of family C G-protein coupled receptors. *Curr. Drug Targets* **8**, 169–184 (2007).
- Kunishima, N. *et al.* Structural basis of glutamate recognition by a dimeric metabotropic glutamate receptor. *Nature* **407**, 971–977 (2000).
- Tsuchiya, D., Kunishima, N., Kamiya, N., Jingami, H. & Morikawa, K. Structural views of the ligand-binding cores of a metabotropic glutamate receptor complexed with an antagonist and both glutamate and Gd^{3+} . *Proc. Natl Acad. Sci. USA* **99**, 2660–2665 (2002).
- Hu, J. & Spiegel, A. M. Structure and function of the human calcium-sensing receptor: insights from natural and engineered mutations and allosteric modulators. *J. Cell. Mol. Med.* **11**, 908–922 (2007).
- Conn, P. J., Christopoulos, A. & Lindsley, C. W. Allosteric modulators of GPCRs: a novel approach for the treatment of CNS disorders. *Nature Rev. Drug Discov.* **8**, 41–54 (2009).
- Serrano-Vega, M. J., Magnani, F., Shibata, Y. & Tate, C. G. Conformational thermostabilization of the $\beta 1$ -adrenergic receptor in a detergent-resistant form. *Proc. Natl Acad. Sci. USA* **105**, 877–882 (2008).
- Shibata, Y. *et al.* Thermostabilization of the neurotensin receptor NTS1. *J. Mol. Biol.* **390**, 262–277 (2009).
- Lebon, G., Bennett, K., Jazayeri, A. & Tate, C. G. Thermostabilisation of an agonist-bound conformation of the human adenosine $\text{A}_{2\text{A}}$ receptor. *J. Mol. Biol.* **409**, 298–310 (2011).
- Gasparini, F. *et al.* [(3H)-M-MPEP, a potent, subtype-selective radioligand for the metabotropic glutamate receptor subtype 5. *Bioorg. Med. Chem. Lett.* **12**, 407–409 (2002).
- Ballesteros, J. A. & Weinstein, H. Integrated methods for the construction of three-dimensional models and computational probing of structure-function relations in G protein-coupled receptors. *Methods Neurosci.* **25**, 366–428 (1995).
- Fredriksson, R., Lagerström, M. C., Lundin, L. G. & Schiöth, H. B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* **63**, 1256–1272 (2003).
- Söding, J. Protein homology detection by HMM-HMM comparison. *Bioinformatics* **21**, 951–960 (2005).
- Goudet, C. *et al.* Heptahelical domain of metabotropic glutamate receptor 5 behaves like rhodopsin-like receptors. *Proc. Natl Acad. Sci. USA* **101**, 378–383 (2004).
- El Moustaine, D. *et al.* Distinct roles of metabotropic glutamate receptor dimerization in agonist activation and G protein coupling. *Proc. Natl Acad. Sci. USA* **109**, 16342–16347 (2012).
- Schwartz, T. W., Frimurer, T. M., Holst, B., Rosenkilde, M. M. & Elling, C. E. Molecular mechanism of 7TM receptor activation—a global toggle switch model. *Annu. Rev. Pharmacol. Toxicol.* **46**, 481–519 (2006).
- Hofmann, K. P. *et al.* A G protein-coupled receptor at work: the rhodopsin model. *Trends Biochem. Sci.* **34**, 540–552 (2009).
- Binet, V. *et al.* Common structural requirements for heptahelical domain function in class A and class C G protein-coupled receptors. *J. Biol. Chem.* **282**, 12154–12163 (2007).
- Chang, W., Chen, T. H., Pratt, S. & Shoback, D. Amino acids in the second and third intracellular loops of the parathyroid Ca^{2+} -sensing receptor mediate efficient coupling to phospholipase C. *J. Biol. Chem.* **275**, 19955–19963 (2000).
- Beqollari, D., Betzenhauser, M. J. & Kammermeier, P. J. Altered G-protein coupling in an mGluR6 point mutant associated with congenital stationary night blindness. *Mol. Pharmacol.* **76**, 992–997 (2009).
- Scheerer, P. *et al.* Crystal structure of opsin in its G-protein-interacting conformation. *Nature* **455**, 497–502 (2008).
- Rasmussen, S. G. F. *et al.* Crystal structure of the $\beta 2$ adrenergic receptor-Gs protein complex. *Nature* **477**, 549–555 (2011).
- Malherbe, P. *et al.* Mutational analysis and molecular modeling of the binding pocket of the metabotropic glutamate 5 receptor negative modulator 2-methyl-6-(phenylethynyl)-pyridine. *Mol. Pharmacol.* **64**, 823–832 (2003).
- Malherbe, P. *et al.* Comparison of the binding pockets of two chemically unrelated allosteric antagonists of the mGlu5 receptor and identification of crucial residues involved in the inverse agonism of MPEP. *J. Neurochem.* **98**, 601–615 (2006).
- Gregory, K. J. *et al.* Probing the metabotropic glutamate receptor 5 (mGlu5) positive allosteric modulator (PAM) binding pocket: discovery of point mutations that engender a “molecular switch” in PAM pharmacology. *Mol. Pharmacol.* **83**, 991–1006 (2013).
- Pagano, A. *et al.* The non-competitive antagonists 2-methyl-6-(phenylethynyl)pyridine and 7-hydroxyiminocyclopropan[b]chromen-1a-carboxylic acid ethyl ester interact with overlapping binding pockets in the transmembrane region of group I metabotropic glutamate receptors. *J. Biol. Chem.* **275**, 33750–33758 (2000).
- O'Brien, J. A. *et al.* A family of highly selective allosteric modulators of the metabotropic glutamate receptor subtype 5. *Mol. Pharmacol.* **64**, 731–740 (2003).
- Ballesteros, J. A., Deupi, X., Olivella, M., Haaksma, E. E. & Pardo, L. Serine and threonine residues bend α -helices in the $\chi_1 = g^-$ conformation. *Biophys. J.* **79**, 2754–2760 (2000).
- Wu, H. *et al.* Structure of a class C GPCR metabotropic glutamate receptor 1 bound to an allosteric modulator. *Science* **344**, 58–64 (2014).

Acknowledgements We thank R. Owen, J. Waterman and D. Axford at I24, Diamond Light Source, Oxford, UK for technical support. We thank C. G. Tate and other colleagues at Heptares Therapeutics Ltd for suggestions and comments, specifically K. Hollenstein and R. Cheng for contributions to the crystallography, and J. Christopher and G. Brown for sourcing allosteric ligands.

Author Contributions J.C.P., G.R.W. and S.K. carried out the conformational thermostabilization of the constructs and determined the stability of the StaR in a panel of reagents/additives to aid purification and crystallization. K.B. carried out the pharmacology. A.J. devised the stabilization strategy and designed the T4L fusion matrix. A.J. and J.C.P. carried out functional analyses the mutants. J.C.E. contributed to construct design and purification strategy. M.S.-V. characterized truncation constructs and established procedures for, and carried out expression and purification. K.O. performed expression, purification, LCP crystallization, optimized purification and performed crystallization in LCP for data collection of the final construct. A.S.D. designed crystallization constructs, established the platform/protocols for, and carried out LCP crystallization and designed crystal optimization, harvested crystals, collected and processed X-ray diffraction data, solved and refined the structure, and devised functional mutations. Computational analysis of the structure and modelling was carried out by B.T. and A.S.D. Project management was carried out by A.J., R.M.C., M.W. and F.H.M. The manuscript was prepared by A.S.D., B.T., A.J., R.M.C., K.B. and F.H.M.

Author Information Co-ordinates and structure factors have been deposited in the Protein Data Bank under the accession code 4009. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.H.M. (fiona.marshall@heptares.com).

METHODS

StaR generation. Full-length human mGlu₅ was used as background for the generation of the conformationally thermostabilized receptor using a mutagenesis approach described earlier^{17,40}. Mutants were analysed for thermostability in the presence of the radioligand [³H]-M-MPEP (Extended Data Fig. 1a). The mGlu₅ StaR contains 6 thermostabilizing mutations (see Extended Data Fig. 1b).

Cell culture. HEK293T cells were cultured in DMEM supplemented with 10% (v/v) fetal bovine serum (FBS). Cells were transfected using GeneJuice (Merck Millipore) according to manufacturer's instructions and collected after 48 h.

Thermostability measurement. Transiently transfected HEK293T cells expressing wild-type receptors were solubilized in 50 mM HEPES pH 7.5, 150 mM NaCl assay buffer containing 1% (w/v) *n*-dodecyl- β -D-maltopyranoside (DDM) and EDTA-free cComplete Protease Inhibitor (Roche) tablets for 1 h rotating at 4 °C. Crude lysates were cleared by centrifugation at 16,000g for 15 min and solubilized receptors purified on Ni-NTA beads. Immobilized receptors were washed in assay buffer containing 0.025% DDM, 20 mM imidazole before elution using assay buffer with 0.025% DDM, 100 mM histidine. Receptor thermostability was measured by incubating with 50 nM [³H]-M-MPEP for 1 h at 4 °C followed by 30 min at varying temperatures. Unbound radioligand was separated by gel filtration and levels of ligand-bound receptor were determined using a liquid scintillation counter. Thermal stability (apparent T_m) was defined as the temperature at which 50% ligand binding is retained.

Truncation and T4-lysozyme fusion constructs. To generate the minimal construct for crystallization, the mGlu₅ StaR construct was truncated at the N terminus up to residue proline 569. In addition, the C terminus of the receptor was truncated at alanine 836. In order to identify the best T4 lysozyme fusion construct, a matrix was designed between Arg 671 and Gln 693 with 67 T4L fusion constructs generated. Following transient expression in HEK293T cells, the fusions were analysed and ranked using thermal stability assay as well as fSEC analysis⁴¹. Based on these data, the best T4L position was found to be between Lys 678 and Lys 679. This final construct is referred to as StaR(569–836)-T4L.

Expression in insect cells. StaR(569–836)-T4L carrying an N-terminal GP64 signal sequence and a C-terminal deca histidine-tag was expressed in Sf21 cells grown in ESF921 medium supplemented with 10% (v/v) FBS and 1% (v/v) penicillin/streptomycin using the FastBac expression system (Invitrogen). Cells were infected at a density of 2×10^6 cells per ml with baculovirus at an approximate multiplicity of infection of 1. Cultures were grown at 27 °C and collected 48 h post-infection.

Membrane preparation and protein purification. All subsequent purification steps were carried out at 4 °C. To prepare membranes, two litres of cells were resuspended in PBS buffer supplemented with protease inhibitor tablets and 5 mM EDTA. Cells were disrupted by micro-fluidizer at 14 kPSI and membranes collected by ultracentrifugation at 204,700g for 1 h. Membranes were washed with PBS buffer supplemented with protease inhibitor tablets and 500 mM NaCl, collected by ultracentrifugation and resuspended in 40 mM HEPES pH 7.5, 250 mM NaCl and stored at –80 °C. Just before solubilization membranes were thawed, homogenized, supplemented with 40 μ M mavoglurant (GVK Bioscience) and 8 mM iodoacetamide, and incubated on a roller mixer for 40 min. Membranes were solubilized with 1.5% (w/v) DDM for one hour, insoluble material was removed by ultra-centrifugation and the solubilized lysate batch bound to 10 ml of Ni-NTA Superflow resin (Qiagen) for three hours in the presence of 10 mM imidazole. Resin was washed with a gradient of 10 to 50 mM imidazole in 40 mM HEPES pH 7.5, 250 mM NaCl, 0.05% (w/v) DDM, and 20 μ M mavoglurant over 35 column volumes before bound material was eluted in a step with 245 mM imidazole. Receptor was further purified by gel filtration (SEC) in 40 mM HEPES pH 7.5, 150 mM NaCl, 0.03% (w/v) DDM, and 40 μ M mavoglurant. Receptor purity was analysed using SDS-PAGE and LC-MS, and receptor monodispersity was assayed by analytical SEC. Protein concentration was determined using the receptor's calculated extinction coefficient at 280 nm ($\epsilon_{280, \text{calc}} = 58,730$ (mg per ml \times cm)^{–1}) and confirmed by quantitative amino acid analysis.

Radioligand binding. HEK293T cells were transiently transfected (as described above) to express either mGlu₅ or mGlu₅StaR(569–836)-T4L. Membranes were prepared as previously described¹⁹. For saturation binding experiments, HEK293T membranes transiently expressing mGlu₅ or mGlu₅StaR(569–836)-T4L (2.5 μ g per well) were incubated with varying concentrations of [³H]-M-MPEP (final assay concentration 0–30 nM) in the presence or absence of 0.1 mM MPEP to define non-specific binding (assay buffer 50 mM HEPES (pH 7.4), 150 mM NaCl). For competition binding experiments, the concentration of [³H]-M-MPEP was fixed to ~1 nM and varying concentrations of cold compound (fenobam, dipragurant, MPEP, mavoglurant; 0.3 nM to 10 μ M) were added to the reaction mixture. Binding assays were incubated for 1 h at 25 °C the reaction was terminated by rapid filtration through 96-well GF/B filter plates pre-soaked with 0.1% polyethylenimine (PEI) using a 96-well head harvester (Tomtec, USA) and plates washed with

5 \times 0.5 ml water. Plates were dried, and bound radioactivity was measured using scintillation spectroscopy on a Microbeta counter (PerkinElmer, UK).

Data were analysed using GraphPad Prism v5 (San Diego, USA). Saturation binding data was globally fitted to one site total and non-specific binding. Inhibition curves were fitted to a four-parameter logistic equation to determine IC₅₀ values, which were converted to K_i values using K_d values determined by saturation binding and the [³H]-M-MPEP concentration of ~1 nM.

Constitutive activity. HEK293 cells were transiently transfected to express wild-type mGlu₅ or mutant constructs using 1 μ g of plasmid DNA per 5 \times 10⁵ cells. Approximately 4 h post-transfection, cells were detached from plate using enzyme free cell dissociation buffer (Life technologies, USA) and total cell count was determined. Cells were seeded into poly-L-lysine treated 96-well plate at the density of 7 \times 10⁴ cells per well and incubated overnight at the 37 °C incubator. Next day, cells were incubated with HBSS buffer supplemented with 5 mM sodium pyruvate and 2 μ l per ml glutamate pyruvate transaminase for 30 min in the 37 °C incubator to breakdown endogenous glutamate. Following 1 h incubation with the IP1 stimulation buffer, the levels of constitutive IP1 were determined using IP-one HTRF assay kit (Cisbio, France) according to the manufacturer's guidelines. In each experiment, sixteen replicates were analysed for each construct. In parallel, cells from the same set of transfections were used to carry out whole cell saturation binding using [³H]-M-MPEP essentially as described above in order to determine receptor numbers under full equilibrium. Levels of IP1 response was then normalized to number of receptors to account for the differences in cell numbers.

Crystallization. mGlu₅StaR(569–836)-T4L was crystallized in LCP at 20 °C. The protein was concentrated to ~25 mg ml^{–1} and mixed with monoolein (Nu-Check) supplemented with 10% (w/w) cholesterol (Sigma Aldrich) and 50 μ M mavoglurant using the twin-syringe method⁴². The final protein:lipid ratio was 40:60 (w/w). 50 nl boli were dispensed on 96-well glass bases and overlaid with 750 nl precipitant solution using a Mosquito LCP from TTPLabtech. 40 μ m plate-shaped crystals of mGlu₅StaR(569–836)-T4L were grown in 100 mM 2-ethanesulphonic acid (MES) across a pH range of 5.5–6.8, 100–250 mM (NH₄)₂HPO₄, 24–34% (v/v) polyethylene glycol 400, and 50 μ M mavoglurant. A complete data set to 2.6 Å was obtained by merging diffraction data from 5 crystals belonging to the monoclinic spacegroup C121. It was possible to mount single crystals for data collection, which were flash-frozen and stored in liquid nitrogen without the addition of further cryoprotectant.

Diffraction data collection and processing. X-ray diffraction data were measured on a Pilatus 6M detector at Diamond Light Source beamline I24 using a beam size of 10 \times 10 μ m. Crystals displayed slightly anisotropic diffraction, initially out to 2.3 Å following exposure to a 70% attenuated beam for 0.5 s per degree of oscillation. It was possible to collect ~30° of useful data from each crystal before radiation damage became severe. Data from 5 individual crystals were integrated using XDS⁴³ and a complete data set compiled using the data collection strategy option of the program Mosflm⁴⁴. Data merging and scaling was carried out using the program AIMLESS⁴⁵. Data collection statistics are reported in Extended Data Table 2.

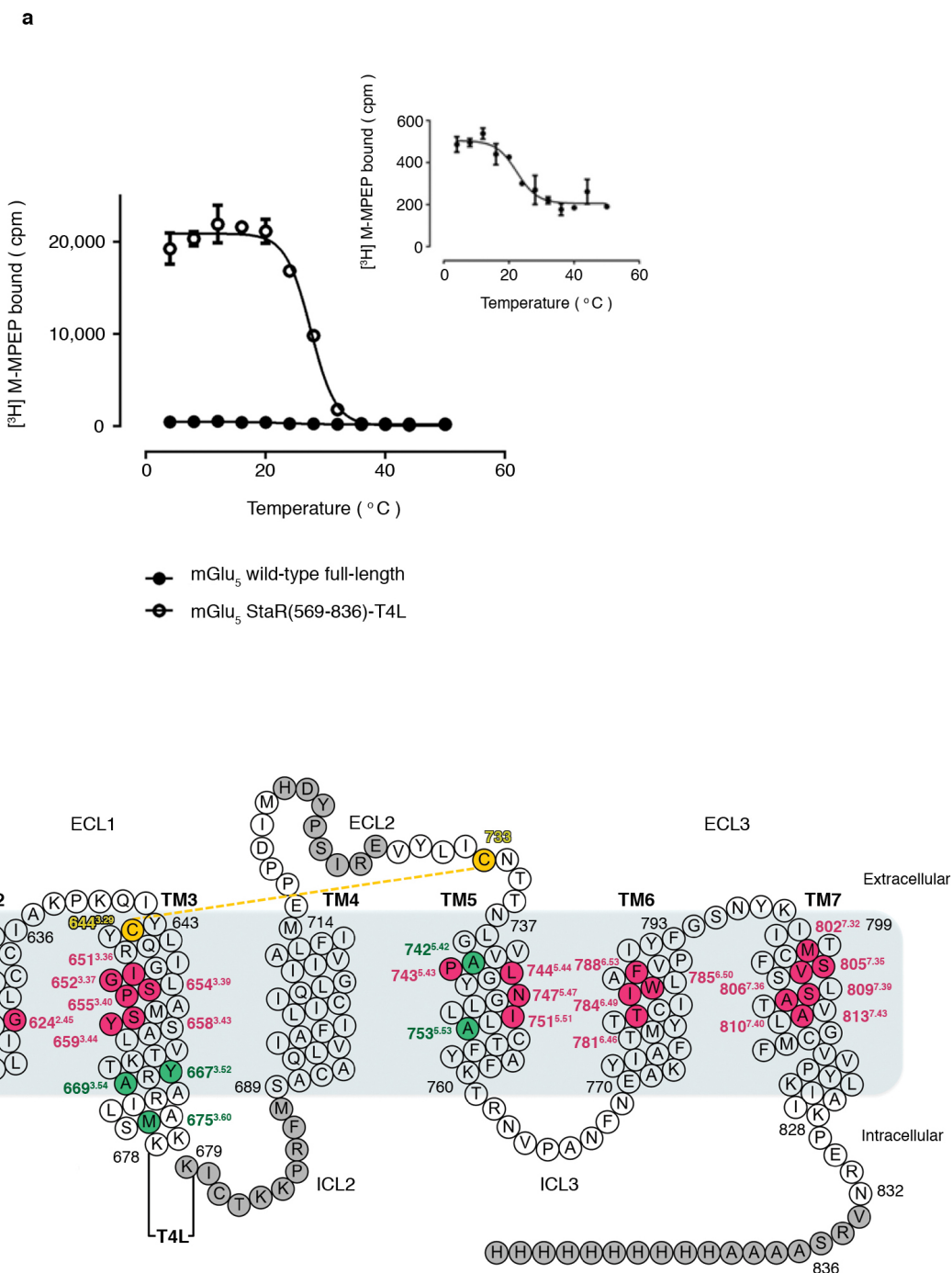
Structure solution and refinement. The structure was solved by molecular replacement using the program Phaser^{45,46} using two independent search models. First T4L from the adenosine A_{2A} receptor structure (PDB ID: 3EML) was located by Phaser, this was then 'fixed' and subsequent searches performed using an in-house library of truncated versions of receptor TMD structures present in the RCSB in attempting to locate the mGlu₅ TMD. However, searches with single receptor structures did not yield a solution for the mGlu₅ TMD that could be validated through subsequent refinement. Finally, a superposed ensemble of eight receptor structures, truncated to a specific and highly sensitive core TM helix region, were required by Phaser to solve the structure and provide an 'averaged' solution model that could be validated in refinement. Initial refinement was carried out with REFMAC5^{44,47}. Manual model building was performed in COOT⁴⁸ using sigma-A weighted 2F_o–F_c, F_o–F_c maps in concert with simulated-annealing and simple composite omit maps calculated in PHENIX⁴⁹ using autobuild. The later stages of refinement were performed using PHENIX implementing a combination of simulated annealing, TLS, positional and individual isotropic B-factor refinement. The structure was refined to 2.6 Å with $R_{\text{work}}/R_{\text{free}} = 23.9/27.4\%$. Structure quality was assessed with MolProbity⁵⁰. Refinement statistics are presented in Extended Data Table 2.

Structure analysis. Structures were superposed and aligned for comparison purposes using the program COOT⁴⁸ to generate global structural superpositions. Local r.m.s.d. analysis between mGlu₅, class B and selected class A GPCRs (Extended Data Table 5) was performed using Maestro v. 9.3 (Schrödinger). Figures were prepared using PyMOL (Schrödinger).

40. Robertson, N. *et al.* The properties of thermostabilised G protein-coupled receptors (StaRs) and their use in drug discovery. *Neuropharmacology* **60**, 36–44 (2011).

41. Kawate, T. & Gouaux, E. Fluorescence-detection size-exclusion chromatography for precrystallization screening of integral membrane proteins. *Structure* **14**, 673–681 (2006).

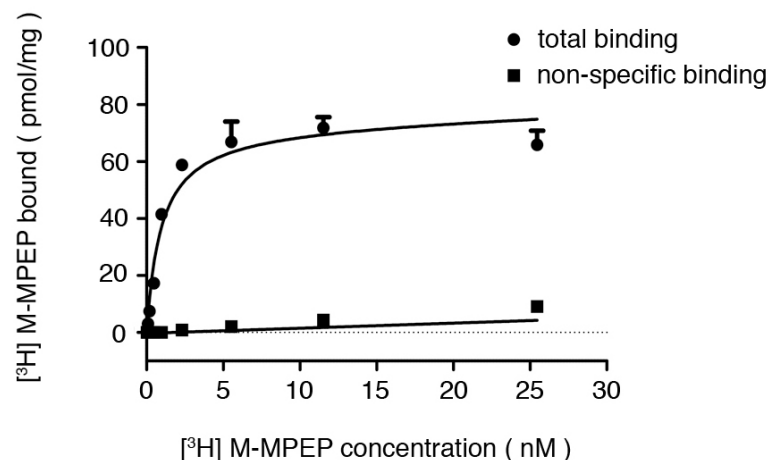
42. Caffrey, M. & Cherezov, V. Crystallizing membrane proteins using lipidic mesophases. *Nature Protocols* **4**, 706–731 (2009).
43. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
44. Leslie, A. G. W. & Powell, H. R. Processing diffraction data with Mosflm. *Evolving Meth. Macromolecular Crystallogr.* **245**, 41–51 (2007).
45. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
46. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
47. Murshudov, G. N. *et al.* REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
48. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
49. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
50. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).



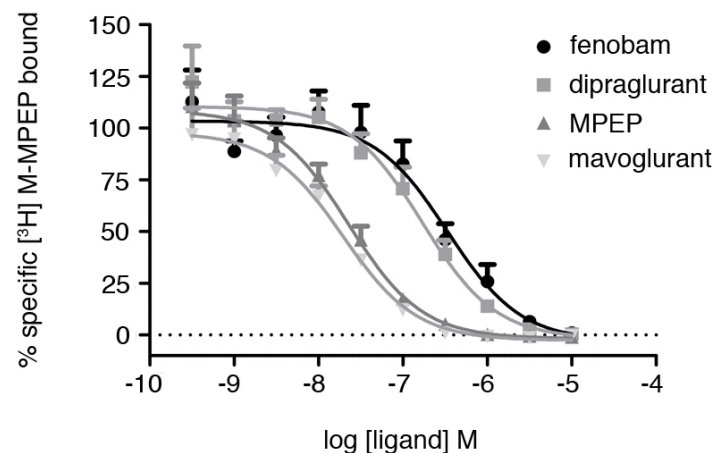
Extended Data Figure 1 | Comparison of wild-type and thermostabilized mGlu₅. **a**, The thermal stability of mGlu₅ constructs measured using [³H] M-MPEP binding following DDM solubilization. Wild-type full-length mGlu₅ (closed circles) has a T_m of 20.6 ± 1.6 °C and mGlu₅StaR 569–836-T4L (open circles) has a T_m of 27.2 ± 0.3 °C. The inset shows the wild-type full-length mGlu₅ data on a different scale. Data represent the mean \pm s.d. from 3

independent experiments. **b**, mGlu₅ crystallization construct (StaR(569–836)-T4L) in schematic representation. Thermostabilizing mutations (green) are: E579A, N667Y, I669A, G675M, T742A, S753A. Residues forming the allosteric pocket are pink. Disordered residues in the structure are grey. The disulphide bond between Cys644^{3,29}–Cys733 is denoted by a dashed yellow line.

a



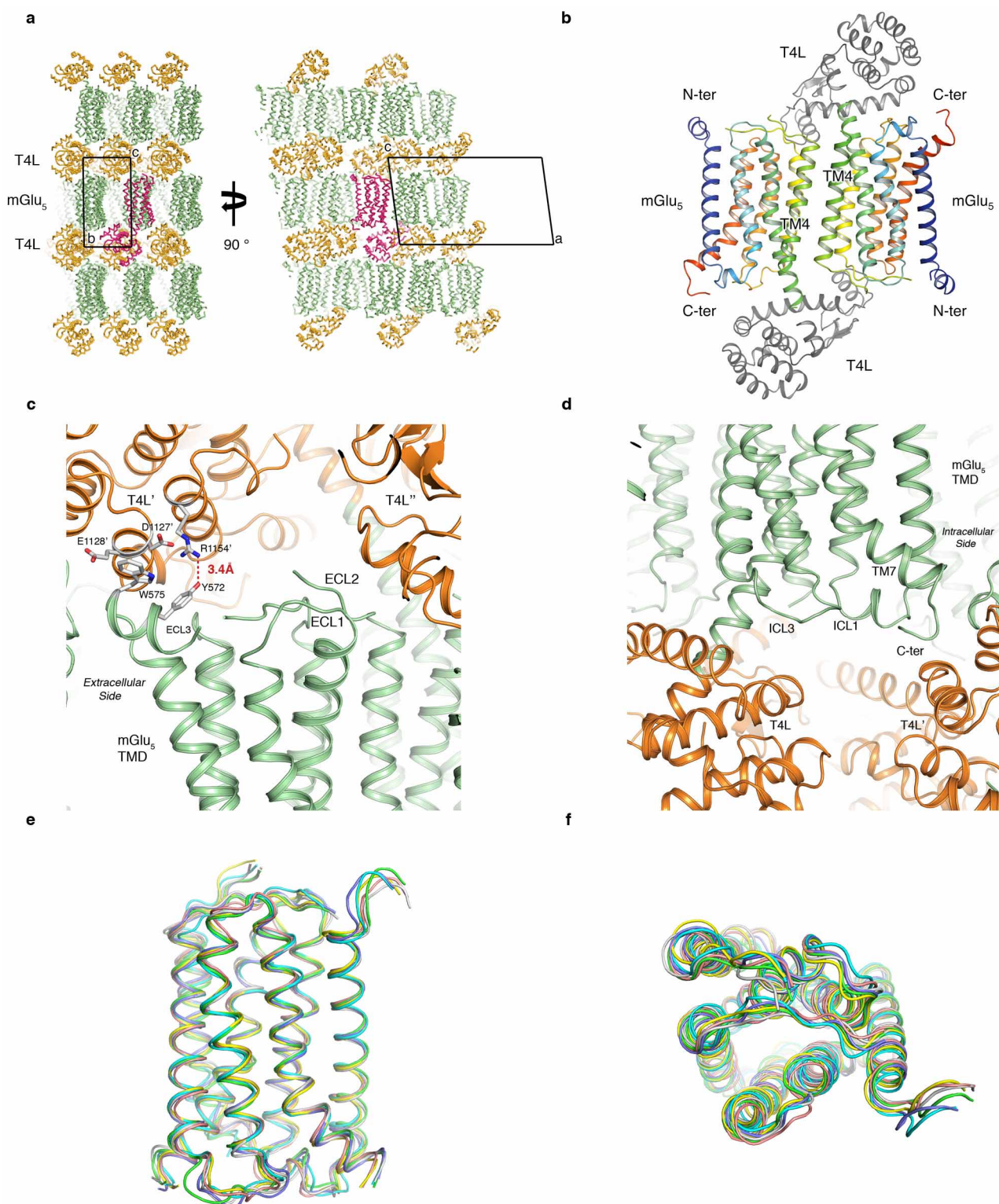
b



Extended Data Figure 2 | Pharmacology of mGlu₅StaR(569–836)-T4L.

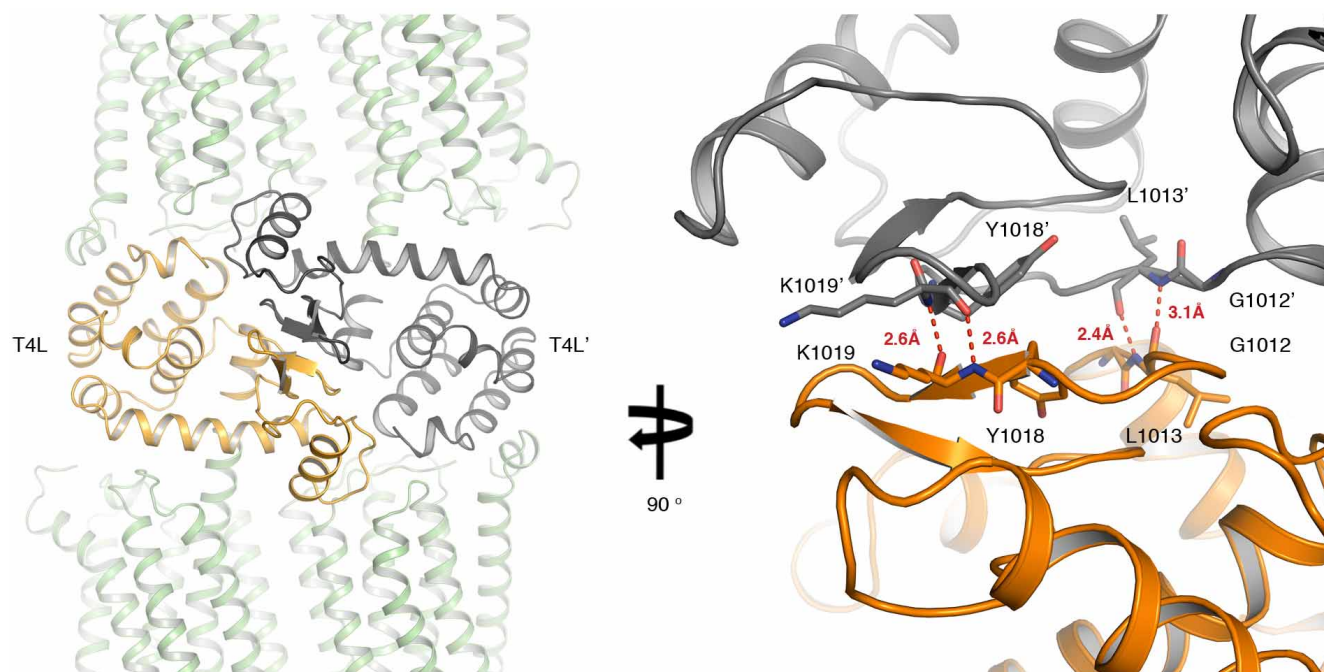
Experiments performed in membranes from HEK293T cells transiently expressing mGlu₅StaR(569–836)-T4L. **a**, Saturation binding of [³H]-M-MPEP to mGlu₅StaR(569–836)-T4L. Non-specific binding was determined by addition of 0.1 mM MPEP. The data shown (mean ± s.e.m.) is representative of three independent experiments performed in duplicate. Data were fitted globally to a one-site saturation isotherm yielding a K_d of 0.86 ± 0.04 nM and B_{max} of 54.7 ± 9.5 pmol mg⁻¹. **b**, Competition binding. Membranes were

incubated with varying concentrations (3 nM–10 μM) of the mGlu₅ NAMs fenobam, dipraglurant, MPEP and mavoglurant. Inhibition curves were fitted to a four-parameter logistic equation to determine IC₅₀ values, which were converted to K_i values using K_d values determined by saturation binding and the [³H]-M-MPEP concentration of ~1 nM. The pK_i values obtained are shown in Extended Data Table 2. Data shown are the mean ± s.e.m. of three independent experiments.



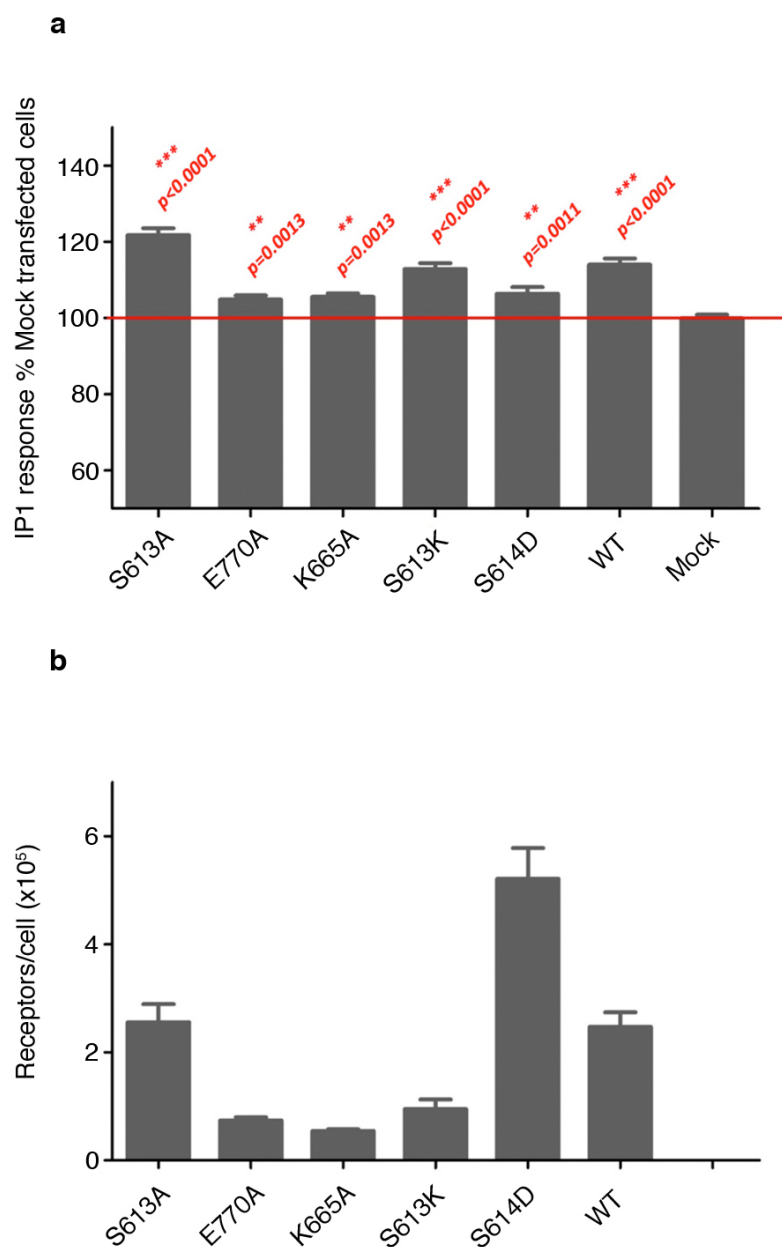
Extended Data Figure 3 | Crystal packing in the mGlu₅StaR(569–836)-T4L monoclinic C121 system. **a**, View of the crystal lattice in the *bc* and *ac* planes respectively. The single copy of the mGlu₅StaR(569–836)-T4L fusion present in the asymmetric unit is shown as C α -trace (magenta). Symmetry mates shown as C α -trace with the receptor TMD coloured green and T4L orange. mGlu₅ receptor TMDs stack along the *c* axis in layers mediated by T4L. **b**, mGlu₅ receptor TMDs pack through an interface mediated by antiparallel TM4–TM4 interactions with the T4L domain swung out towards one side of

the receptor. **c**, The extracellular loops of the mGlu₅ TMD are not constrained by packing interactions. Receptor and T4L coloured as in **a**. Two specific interactions are observed between the mGlu₅-TMD N-terminal helical extension and T4L moiety from a symmetry mate. **d**, The intracellular loops of the mGlu₅ TMD are not constrained by packing interactions, receptor and T4L coloured as in **a**. **e, f**, Result of 10 ns MD simulation on the mGlu₅TMD, models shown are separated by 2 ns of simulation. C α r.m.s.d. (between starting and final model) = 2.0 Å.



Extended Data Figure 4 | T4L mediated contacts in the mGlu₅StaR(569–836)-T4L monoclinic C121 system. T4L layers are held together by extensive contacts between the T4L moieties. Specifically the N-terminal lobes of T4L (residues 1002–1071) are shifted by up to 3.7 Å (distance between equivalent Ca positions of E1022) in comparison to the input structure to molecular replacement and refinement (residues 1002–1071 from PDB ID: 3EML). The

N-terminal lobes of T4L appear to pivot at V1071 to accommodate packing interactions in the mGlu₅StaR(569–836)-T4L monoclinic C121 system, forming an intramolecular β -sheet between residues Y1018–K1019 and Y1018'–K1019' of a symmetry mate, with additional hydrogen bonds between the back bone of G1012–L1013 and G1012'–L1013'. Hydrogen bonds denoted by dashed red lines, distances measured in angstroms.



Extended Data Figure 5 | Constitutive activity of mGlu₅ mutants. IP1 response and number of receptors per cells in HEK293 cells transiently transfected to express wild-type (WT) or the indicated mGlu₅ mutants. **a**, Constitutive activity of wild-type and mutant constructs expressed as percentage of background IP1 levels detected in mock transfected control

sample. **b**, Numbers of expressed receptors per cell were determined using [³H]-M-MPEP binding. Error bars indicate standard error of mean and *P* values are derived from an unpaired two-tailed *t*-test. Data are representative of three independent experiments.

Extended Data Table 1 | Comparison of the pharmacology of mGlu₅ and mGlu₅ StaR(569–836)-T4L**a**

Construct	K_d (nM) \pm S.E.M.	B_{max} (pmol/mg) \pm S.E.M.
mGlu ₅ wild-type full-length	1.05 \pm 0.15	6.1 \pm 0.6
mGlu ₅ StaR(569-836)-T4L	0.86 \pm 0.04	54.7 \pm 9.5 **

(** $p < 0.01$).**b**

Compound	mGlu ₅ wild-type full-length $pK_i \pm$ S.E.M.	mGlu ₅ StaR(569-836)-T4L $pK_i \pm$ S.E.M.
fenobam	7.42 \pm 0.07	6.96 \pm 0.24
dipraglurant	7.49 \pm 0.30	7.24 \pm 0.23
MPEP	8.46 \pm 0.07	8.10 \pm 0.12
mavoglurant	8.25 \pm 0.12	8.17 \pm 0.13

Binding analysis was performed in HEK293 membranes transiently transfected to express either mGlu₅ or mGlu₅StaR(569–836)-T4L. **a**, Following saturation analysis the affinity (K_d) and expression levels (B_{max}) obtained were compared using an unpaired two-tailed *t*-test. Data from three independent experiments performed in duplicate. **b**, Competition binding analysis was carried out to compare the affinity of a range of ligands. Inhibition curves were fitted to a four-parameter logistic equation to determine IC_{50} values, which were converted to K_i values using K_d values determined by saturation binding and the [³H]-M-MPEP concentration of ~1 nM. The pK_i values obtained for each compound binding to either mGlu₅ or mGlu₅StaR(569–836)-T4L were compared using an unpaired two-tailed *t*-test. For each compound there was no significant change in affinity at mGlu₅StaR(569–836)-T4L compared to mGlu₅. Data shown are $pK_i \pm$ s.e.m. for three independent experiments.

Extended Data Table 2 | Data collection and refinement statistics for mGlu₅StaR(569–836)-T4L

Data collection	
Number of crystals	5
Space group	<i>C121</i>
Cell dimensions	
a, b, c (Å)	143.2, 43.6, 82.0
α , β , γ (°)	90.0, 99.4, 90.0
Number of reflections measured	39243
Number of unique reflections	14800
Resolution (Å)	34.51 – 2.60 (2.72 – 2.60)
R_{merge}	0.116 (0.754)
Mean $I/\sigma(I)$	7.4 (1.8)
Completeness (%)	94.9 (94.8)
Redundancy	2.7 (2.6)
Refinement	
Resolution (Å)	29.74 – 2.60
Number of reflections (test set)	14093 (691)
$R_{\text{work}}/R_{\text{free}}$	0.2392 / 0.2747
Number of atoms	
All	3355
Proteins	3212
Ligand	23
Others (Lipids, ions, waters)	120
Average B factors (Å ²)	
All	41.49
mGlu ₅	39.25
T4L lysozyme	44.22
Ligand	34.58
Others (Lipid, ion, water)	49.22
RMSD	
Bond lengths (Å)	0.003
Bond angles (°)	0.727
Ramachandran statistics	
Favored regions (%)	98.0
Allowed regions (%)	2.0
Outliers (%)	0.0
<i>MolProbity</i> overall score (percentile)	1.17 (100 th percentile)

Values in parenthesis indicate highest resolution shell.

Extended Data Table 3 | Representative alignment of 22 human class C receptor sequences (taken from an alignment of 156 class C receptor sequences)

	1.43 - 583	1.44 - 584	1.45 - 585	1.46 - 586	1.47 - 587	1.48 - 588	1.49 - 589	1.50 - 590	1.51 - 591	1.52 - 592	1.53 - 593	1.54 - 594	1.55 - 595	1.56 - 596	1.57 - 597	2.43 - 622	2.44 - 623	2.45 - 624	2.46 - 625	2.47 - 626	2.48 - 627	2.49 - 628	2.50 - 629	2.51 - 630	2.52 - 631	2.53 - 632	2.54 - 633	2.55 - 634	2.56 - 635	2.57 - 636	3.43 - 658	3.44 - 659	3.45 - 660	3.46 - 661	3.47 - 662	3.48 - 663	3.49 - 664	3.50 - 665	3.51 - 666	3.52 - 667	3.53 - 668	3.54 - 669	3.55 - 670	3.56 - 671	3.57 - 672											
mGlu5 (human)	A	V	A	F	A	C	L	G	I	L	A	T	L	F	V	L	A	G	I	C	L	G	V	L	C	T	F	F	C	L	I	S	Y	S	A	L	V	T	K	T	N	R	I	A	R	I										
mGlu1 (human)	A	V	A	F	A	C	L	G	I	L	A	T	L	F	V	L	A	G	I	C	L	G	V	L	C	T	F	F	C	L	I	S	Y	S	A	L	V	T	K	T	N	R	I	A	R	I										
mGlu2 (human)	P	V	T	I	A	C	L	G	A	L	A	T	L	F	V	L	A	G	I	C	L	G	V	L	C	T	F	F	C	L	I	S	Y	S	A	L	V	T	K	T	N	R	I	A	R	I										
mGlu3 (human)	P	V	T	I	A	C	L	G	A	L	A	T	L	F	V	L	A	G	I	C	L	G	V	L	C	T	F	F	C	L	I	S	Y	S	A	L	V	T	K	T	N	R	I	A	R	I										
mGlu4 (human)	P	L	L	L	A	V	V	G	I	A	A	T	L	F	V	L	A	G	I	C	L	G	V	L	C	T	F	F	C	L	I	S	Y	S	A	L	V	T	K	T	N	R	I	A	R	I										
mGlu6 (human)	P	L	L	L	A	V	V	G	I	V	A	T	T	T	V	L	A	G	I	C	L	G	V	L	C	T	F	F	C	L	I	S	Y	S	A	L	V	T	K	T	N	R	I	A	R	I										
mGlu7 (human)	P	V	F	V	A	M	L	G	I	I	A	T	I	F	V	L	A	G	I	C	L	G	V	L	C	T	F	F	C	L	I	S	Y	S	A	L	V	T	K	T	N	R	I	A	R	I										
mGlu8 (human)	P	V	F	V	A	I	L	G	I	I	A	T	T	F	V	L	A	G	I	C	L	G	V	L	C	T	F	F	C	L	I	S	Y	S	A	L	V	T	K	T	N	R	I	A	R	I										
CaSR (human)	L	T	S	F	A	V	L	G	I	F	L	A	A	F	V	L	A	G	I	C	L	G	V	L	C	T	F	F	C	L	I	S	Y	S	A	L	V	T	K	T	N	R	I	A	R	I										
GABA1 (human)	V	S	V	L	S	S	L	G	I	V	L	A	V	V	C	L	A	G	I	C	L	G	V	L	C	T	F	F	C	L	I	S	Y	S	A	L	V	T	K	T	N	R	I	A	R	I										
GABA2 (human)	L	S	A	F	T	I	C	G	M	I	M	A	S	A	F	L	A	G	I	C	L	G	V	L	C	T	F	F	C	L	I	S	Y	S	A	L	V	T	K	T	N	R	I	A	R	I										
GPR156 (human)	V	W	T	A	L	T	S	C	G	L	M	L	I	L	F	F	L	A	G	I	C	L	G	V	L	C	T	F	F	C	L	I	S	Y	S	A	L	V	T	K	T	N	R	I	A	R	I									
GPR158 (human)	I	I	S	F	Q	A	C	C	M	L	L	D	F	V	S	L	A	G	I	C	L	G	V	L	C	T	F	F	C	L	I	S	Y	S	A	L	V	T	K	T	N	R	I	A	R	I										
GPR179 (human)	V	L	A	C	Q	A	C	C	M	L	L	A	I	F	L	S	L	A	G	I	C	L	G	V	L	C	T	F	F	C	L	I	S	Y	S	A	L	V	T	K	T	N	R	I	A	R	I									
TAS1R1 (human)	L	L	A	A	N	T	L	L	L	L	L	L	G	T	L	L	G	S	V	G	S	L	V	P	Y	V	F	L	F	C	I	S	C	I	A	V	R	R	S	F	Q	I	L	I	A	R	I									
TAS1R2 (human)	V	A	L	L	A	A	L	L	G	F	L	S	T	L	A	I	L	V	C	M	V	V	P	Y	V	F	L	F	P	C	I	S	C	I	A	V	R	R	S	F	Q	I	L	I	A	R	I									
TAS1R3 (human)	L	L	L	L	L	S	L	A	L	G	L	V	L	A	A	L	V	C	M	V	V	P	Y	V	F	L	F	P	C	I	S	C	I	A	V	R	R	S	F	Q	I	L	I	A	R	I										
GPRC5a (human)	L	E	T	V	A	T	A	G	V	V	T	S	V	A	F	L	L	G	V	T	L	G	L	F	F	I	C	F	S	C	L	L	L	A	H	Q	A	W	S	L	V	R	R	I	A	R	I									
GPRC5b (human)	V	E	A	V	A	G	A	G	V	A	L	I	T	L	L	L	L	G	V	T	L	G	L	F	F	I	C	F	S	C	L	L	L	A	H	Q	A	W	S	L	V	R	R	I	A	R	I									
GPRC5c (human)	L	E	A	V	A	G	A	I	V	T	T	F	V	L	L	L	L	G	V	T	L	G	L	F	F	I	C	F	S	C	L	L	L	A	H	Q	A	W	S	L	V	R	R	I	A	R	I									
GPRC5d (human)	L	E	S	L	A	I	L	G	I	V	V	T	I	L	L	L	L	G	V	T	L	G	L	F	F	I	C	F	S	C	L	L	L	A	H	Q	A	W	S	L	V	R	R	I	A	R	I									
GPC6A (human)	L	L	I	L	S	L	L	G	I	I	V	F	V	L	V	L	L	C	H	F	L	N	F	A	S	A	T	S	F	F	I	C	I	S	C	I	T	K	S	L	K	I	L	A	R	I										
TM1															TM2															TM3																										
	4.39 - 700	4.40 - 701	4.41 - 702	4.42 - 703	4.43 - 704	4.44 - 705	4.45 - 706	4.46 - 707	4.47 - 708	4.48 - 709	4.49 - 710	4.50 - 711	4.51 - 712	4.52 - 713	4.53 - 714	5.43 - 743	5.44 - 744	5.45 - 745	5.46 - 746	5.47 - 747	5.48 - 748	5.49 - 749	5.50 - 750	5.51 - 751	5.52 - 752	5.53 - 753	5.54 - 754	5.55 - 755	5.56 - 756	5.57 - 757	6.43 - 778	6.44 - 779	6.45 - 780	6.46 - 781	6.47 - 782	6.48 - 783	6.49 - 784	6.50 - 785	6.51 - 786	6.52 - 787	6.53 - 788	6.54 - 789	6.55 - 790	6.57 - 792												
mGlu5 (human)	L	I	C	I	Q	L	G	I	I	V	A	L	F	I	M'	P	L	G	Y	N	G	L	L	I	L	S	C	T	F	Y	M	Y	T	T	C	I	I	W	L	A	F	V	P	I	Y											
mGlu1 (human)	L	I	S	V	Q	L	L	I	V	V	A	L	I	I	M	P	L	G	Y	N	V	L	L	I	I	M	S	C	T	L	Y	M	Y	T	T	C	I	I	W	L	A	F	V	P	I	Y										
mGlu2 (human)	L	I	S	G	Q	L	L	I	V	V	A	W	L	I	V	S	L	T	Y	D	V	I	L	V	I	L	C	T	L	Y	M	Y	T	T	C	I	I	W	L	A	F	V	P	I	Y											
mGlu3 (human)	L	I	L	V	Q	L	I	V	M	S	V	W	L	I	L	L	L	T	G	Y	S	M	L	L	M	V	T	C	T	V	Y	M	Y	T	T	C	I	I	W	L	A	F	V	P	I	Y										
mGlu4 (human)	L	I	S	L	Q	L	L	G	I	C	V	W	F	V	V	L	L	T	G	Y	S	M	L	L	M	V	T	C	T	V	Y	M	Y	T	T	C	I	I	W	L	A	F	V	P	I	Y										
mGlu6 (human)	L	T	S	L	Q	V	V	G	M	I	A	W	L	F	G	A	C	L	G	Y	S	I	L	L	M	M	V	T	C	T	V	Y	M	Y	T	T	C	I	I	W	L	A	F	V	P	I	Y									
mGlu7 (human)	L	I	S	V	Q	L	L	G	V	F	I	W	F	G	V	L	L	T	G	Y	S	I	L	L	M	M	V	T	C	T	V	Y	M	Y	T	T	C	I	I	W	L	A	F	V	P	I	Y									
mGlu8 (human)	L	I	S	V	Q	L	L	G	V	V	F	W	F	V	V	L	L	T	G	Y	S	I	L	L	M	M	V	T	C	T	V	Y	M	Y	T	T	C	I	I	W	L	A	F	V	P	I	Y									
CaSR (human)	C	T	F	M	Q	L	I	V	I	C	V	I	W	L	Y	T	F	Y	A	Y	K	G	L	L	L	M	L	L	G	I	F	L	I	Y	N	V	A	V	L	C	C	I	I	A	P	V	T	S								
GABA1 (human)	L	V	G	M	D	V	L	T	L	A	I	W	Q	A	I	V	V	Y	A	Y	K	G	L	L	L	M	L	L	F	G	C	F	L	V	N	L	V	L	A	F	H	T	A	R	F	V	T	L								
GABA2 (human)	M	L	L	I	D	L	C	I	L	I	C	W	Q	A	I	V	V	A	E	L	L	F	L	L	L	W	G	V	Y	L	C	Y	E	L	L	I	S	A	F	H	T	A	R	F	V	T	L									
GPR156 (human)	R	M	L	M	A	D	V	I	L	L	M	T	W	F	L	I	V	A	E	L	L	F	L	L	C	W	G	S	F	L	C	Y	E	L	L	I	S	A	F	H	T	A	R	F	V	T	L									
GPR158 (human)	R	S	L	A	Q	L	L	I	C	V	L	T	W	L	V	V	L	V	A	E	L	L	N	G	L	L	S	I	S	A	F	A	C	E	L	L	F	N	F	V	S	S	A	F	H	T	A	R	F	V	T	L				
GPR179 (human)	S	S	A	G	A	Q	L	L	I	C	V	L	T	W	L	V	V	L	N	T	S	L	D	A	T	L	L	S	V	V	G	F	S	F	M	L	F	N	F	V	T	S	S	V	S	A	F	H	T	A	R	F	V	T	L	
TAS1R1 (human)	R	S	L	A	Q	L	L	I	C	V	L	T	W	L	V	V	L	V	A	E	L	L	N	G	L	L	L	S	I	S	A	F	A	C	E	L	L	F	N	F	V	T	S	S	V	S	A	F	H	T	A	R	F	V	T	L
TAS1R2 (human)	I	T	V	L	K	M	V	I	V	T	I	G	M	L	A	A	L	H	A	T	N	A	D	T	L	L	S	A	F	L	G	F	L	G	M	L	A	Y	F	I	T	W	V	S	F	V	P	L	M	Y						
TAS1R3 (human)	A	M	L	V	E	V	A	L	C	T	W	Y	L	V	L	L	L	A	L	I	Y	D	M	V	L	L	L	L	V	T	L	G	L	M	L	A	Y	F	I	T	W	V	S	F	V	P	L	M	Y							
GPRC5a (human)	F	S	L	V	Q	D	V	I	A	I	E	Y	I	V	L	L	L	A	L	I	Y	D	M	V	L	L	L	L	V	T	L	G	L	M	L	A	Y	F	I	T	W	V	S	F	V	P	L	M	Y							
GPRC5b (human)	L	M	L	V	Q	V	I	I	A	V	E	W	L	V	L	L	L	A	L	I	Y	D	M	V	L	L	L	L	V	T	L	G	L	M	L	A	Y	F	I	T	W	V	S	F	V	P	L	M	Y							
GPRC5c (human)	L	T	L	V	E	V	I	I	N	T	E	W	L	V	L	I	L	A	L	I	Y	V	M	L	L	L	L	L	G	A	F	L	G	L	A	T	S	V	I	I	W	V	W	I	S	M	L	Y								
GPRC5d (human)	C	S	L	Q	Q	I	I	A	T	E	Y	V	T	L	L	L	L	M	L	V	Y	M	L	L	L	L	L	L	T	F	F	I	V	V	L	F	S	I	I	W	V	W	I	S	M	L	Y									
GPC6A (human)	C	T	G	I	Q	V	I	C	T	A	L	L	I	F	F	L	L	G	Y	I	A	I	L	L	A	F	I	C																												

Numbers at top of alignment refer to residue numbers from mGlu₅. Numbers in superscript also across the top of each TM alignment are given to amino acid residues and based on a modification of the Ballesteros-Weinstein numbering system suggested for class C in ref. 1 with X.50 residues denoted by black columns. TM alignments cover core 15 residues of each TM helix as observed in the mGlu₅ crystal structure. Sequences from top to bottom of each alignment include; 8 metabotropic glutamate receptors; calcium sensing receptor (CaS); γ -aminobutyric acid_B receptors (GABA_B); GPR156/158/179; TAS1R1-3 taste receptors; retinoic acid-induced G1-G4; GPC6A. The pink highlighted column present in each TM alignment denotes the X.50 Ballesteros-Weinstein²¹ residue defined in ref. 39 on the basis of the alignment of mGlu₁ and CXCR4. Note X.50 numbering of both systems is in agreement across TM5.

Extended Data Table 4 | Structural differences between mGlu₅, class B and selected class A GPCRs

Receptor	PDB ID	RMSD							
		All TMs	TM1	TM2	TM3	TM4	TM5	TM6	TM7
CRF ₁	4K5Y	3.6	2.0(0.5)	2.3(2.0)	2.9(1.6)	3.8(1.6)	2.4(0.6)	3.1(1.3)	6.9(2.5)
Rhodopsin	1F88	3.3	2.0(0.9)	3.7(2.9)	2.6(1.5)	3.5(1.0)	4.7(1.9)	2.5(1.5)	3.9(2.9)
β_1 adrenergic	2VT4	3.2	2.5(0.8)	3.4(3.0)	2.2(1.5)	3.2(1.3)	3.6(1.9)	2.5(1.5)	4.5(3.0)
β_2 adrenergic	2RH1	3.1	2.4(0.7)	3.5(3.0)	2.2(1.5)	3.0(1.0)	3.5(2.0)	2.6(1.6)	4.4(3.1)
β_2 adrenergic	3SN6	3.5	2.8(1.0)	3.4(3.0)	2.8(1.4)	2.6(1.2)	3.8(1.9)	5.1(2.4)	3.6(2.3)
Adenosine A _{2A}	2YDV	3.2	2.0(0.4)	3.4(3.0)	3.6(1.6)	2.9(0.9)	3.7(2.0)	2.8(1.6)	3.7(2.1)
Adenosine A _{2A}	3PWH	3.1	1.5(0.9)	3.2(3.0)	2.8(1.8)	3.2(1.3)	3.9(2.1)	2.4(1.5)	4.4(2.9)
CXCR4	3ODU	3.4	2.3(1.4)	3.1(2.7)	2.4(1.5)	3.5(2.3)	4.4(2.0)	3.1(1.5)	4.8(3.0)
Dopamine D ₃	3PBL	3.1	1.7(0.7)	3.2(2.9)	2.6(1.4)	2.9(1.0)	3.7(1.9)	2.5(1.4)	4.5(2.8)
Histamine H ₁	3RZE	3.1	1.9(0.7)	3.3(2.9)	2.2(1.5)	2.9(1.3)	4.2(1.8)	2.6(1.7)	4.0(2.8)
Muscarinic M ₂	3UON	3.3	2.1(0.8)	3.5(2.9)	1.9(1.4)	3.5(1.2)	4.5(1.6)	2.9(1.4)	4.1(2.6)
Muscarinic M ₃	4DAJ	3.1	1.8(0.6)	3.6(2.9)	2.1(1.6)	3.2(1.1)	4.3(1.7)	3.0(2.1)	3.6(2.6)
S1P ₁	3V2Y	2.9	2.1(0.6)	2.5(1.9)	2.7(1.9)	3.2(1.0)	3.0(0.7)	2.1(1.4)	4.1(3.0)
κ opioid	4DJH	3.4	3.6(1.1)	2.9(2.3)	2.7(1.8)	3.5(1.5)	4.5(2.1)	2.7(1.4)	3.9(2.3)
μ opioid	4DKL	3.4	2.3(0.9)	2.8(2.3)	2.7(1.8)	3.1(1.4)	4.8(2.0)	2.8(1.5)	4.7(2.9)
δ opioid	4EJ4	3.4	2.5(0.8)	3.0(2.6)	2.5(1.7)	3.4(1.3)	4.6(2.0)	2.7(1.6)	4.7(2.9)
ORL-1	4EA3	3.2	2.1(1.0)	3.0(2.5)	2.4(1.6)	3.2(1.3)	4.6(1.9)	2.6(1.5)	4.5(2.9)
PAR-1	3VW7	3.5	3.0(1.5)	3.2(2.5)	3.5(1.8)	2.8(1.1)	4.6(0.9)	3.5(2.0)	3.7(3.0)
NTSR ₁	4GRV	3.3	3.4(0.9)	3.1(2.7)	2.9(1.3)	3.0(1.1)	3.7(2.0)	3.6(1.4)	3.3(2.0)

Backbone r.m.s.d. values were calculated after a global superposition using a core TM region shared by class A and C GPCRs and mGlu₅ as defined by mGlu₅ residues 579–603, 613–638, 642–674, 692–712, 740–761, 770–790 and 797–819, corresponding to the class A Ballesteros-Weinstein residues 1.35–1.59, 2.38–2.63, 3.23–3.55, 4.41–4.61, 5.40–5.61, 6.33–6.53, and 7.33–7.55. Backbone r.m.s.d. values in brackets were calculated after local superposition of the individual TM helices. Chain C was used for CRF₁R and chain A for the class A GPCRs.

Extended Data Table 5 | Alignment and comparison of allosteric binding site residues across the 8 human metabotropic glutamate receptors

PIN	mGlu ₅	mGlu ₅	mGlu ₁	mGlu ₂	mGlu ₃	mGlu ₄	mGlu ₆	mGlu ₇	mGlu ₈	mGlu*	Conservation	quality**	Consensus
2.45	624	G	G	G	G	G	G	G	G	G	11	17	100
2.46	625	I	I	V	V	I	I	I	I	I	9	16.7	77
2.49	628	G	G	C	S	C	I	C	C	C	5	7.4	55
3.36	651	I	V	L	L	L	L	L	L	I	9	16.2	66
3.37	652	G	G	G	G	G	G	G	G	G	11	17	100
3.39	654	S	S	A	S	G	G	G	G	G	8	13.9	55
3.40	655	P	S	F	F	M	T	M	M	F	4	8.4	33
3.43	658	S	C	C	C	S	S	S	S	I	5	9.4	55
3.44	659	Y	Y	Y	Y	Y	Y	Y	Y	Y	11	17	100
5.43	743	P	P	S	S	L	C	S	S	S	4	10	55
5.44	744	L	L	L	L	L	L	L	L	Q	7	13.2	88
5.47	747	N	N	N	D	S	S	S	S	N	7	13.9	44
5.51	751	I	I	I	V	M	M	M	M	I	8	15.2	44
6.46	781	T	T	T	T	T	T	T	T	T	11	17	100
6.49	784	I	I	I	I	V	I	V	I	I	9	16.4	77
6.50	785	W	W	W	W	W	W	W	W	W	11	17	100
6.53	788	F	F	F	F	F	F	F	F	F	11	17	100
7.32	802	M	T	M	M	L	L	L	L	L	6	14.6	55
7.35	805	S	A	S	S	S	S	S	S	S	8	16.4	88
7.36	806	V	V	V	V	V	L	M	M	I	8	15.5	55
7.39	809	S	S	S	S	S	S	S	S	S	11	17	100
7.40	810	A	V	G	G	A	A	A	A	A	8	13.4	66
7.43	813	A	A	V	V	S	S	A	S	A	6	13.1	44

All sequences are from *Homo sapiens* apart from column 11 which is from *Drosophila melanogaster*. Letters in columns 3–11 are single letter amino acid code. Columns are from left: (1) PIN number based on a modification of the Ballesteros-Weinstein numbering system suggested for class C in ref. 1; (2) mGlu₅ residue number; (3) mGlu₅ residue, single letter code; (4) mGlu₁ equivalent residue, single letter code; (5) mGlu₂ equivalent residue, single letter code; (6) mGlu₃ equivalent residue, single letter code; (7) mGlu₄ equivalent residue, single letter code; (8) mGlu₆ equivalent residue, single letter code; (9) mGlu₇ equivalent residue, single letter code; (10) mGlu₈ equivalent residue, single letter code; (11) mGlu* equivalent residue, single letter code; (12) overall residue conservation at equivalent positions; (13) overall residue quality score at equivalent positions**. The quality score is inversely proportional to the average cost of all pairs of mutations observed in a particular column of the alignment, a high alignment quality score for a column would suggest that there are no mutations, or most mutations observed are favourable; (14) overall alignment consensus.

Velocity anti-correlation of diametrically opposed galaxy satellites in the low-redshift Universe

Neil G. Ibata¹, Rodrigo A. Ibata², Benoit Famaey² & Geraint F. Lewis³

Recent work has shown that the Milky Way and the Andromeda galaxies both possess the unexpected property that their dwarf satellite galaxies are aligned in thin and kinematically coherent planar structures^{1–7}. It is interesting to evaluate the incidence of such planar structures in the larger galactic population, because the Local Group may not be a representative environment. Here we report measurements of the velocities of pairs of diametrically opposed satellite galaxies. In the local Universe (redshift $z < 0.05$), we find that satellite pairs out to a distance of 150 kiloparsecs from the galactic centre are preferentially anti-correlated in their velocities (99.994 per cent confidence level), and that the distribution of galaxies in the larger-scale environment (out to distances of about 2 megaparsecs) is strongly clumped along the axis joining the inner satellite pair ($>7\sigma$ confidence). This may indicate that planes of co-rotating satellites, similar to those seen around the Andromeda galaxy, are ubiquitous, and their coherent motion suggests that they represent a substantial repository of angular momentum on scales of about 100 kiloparsecs.

The satellite galaxies of the Milky Way have long been known to be preferentially located close to a plane⁸, but this observation could be dismissed as a mere coincidence. However, as faint galaxies were uncovered in the Sloan Digital Sky Survey⁹ (SDSS), it became clear that our Galaxy hosts a planar structure of satellites with a close-to-polar orientation^{10–12}. The complications due to spatial incompleteness of satellite samples that complicate analyses in the Milky Way are largely alleviated when observing the next-nearest giant galaxy, Andromeda (M31). The presence of a vast plane of co-rotating dwarf galaxies was recently detected in that galaxy as a result of new photometric^{13–15} and spectroscopic^{16,17} surveys of its halo. A full 50% of the dwarf galaxies around M31 belong to this structure^{5,6}. The satellites in the plane have galactocentric distances as great as ~ 300 kpc, yet they display very small scatter (12.6 kpc) in the direction perpendicular to the plane, and they possess coherent kinematics, suggestive of common rotation about their host. Recent analyses have also uncovered possible galaxy alignments in the M81 system¹⁸ and in the NGC 3109 association¹⁹. Such satellite alignments may arise naturally if dwarf galaxies formed from tidal debris left over from ancient galaxy mergers^{2,20}, but this scenario remains difficult to reconcile with the high dark matter content deduced for these objects²¹. Although the presence of the planar structures in the Local Group is now firmly established, they may represent a fossil of the particular dynamical formation history of the Milky Way and Andromeda systems²². It is therefore necessary to investigate more distant systems to ascertain the true significance of these local detections.

We devised a test (Methods) to quantify the incidence of planar systems of satellites. Beyond a few megaparsecs from Earth, reliable and accurate relative distance measurements are beyond our present technological capabilities; this means that we have to deal with two-dimensional projections of galactic systems, possessing only the radial component of velocity. We take the M31 system as a template for the search for satellite alignments, because its global structure and dynamics are at present the best understood. Half of that system shows coherent rotation, which means that for orientations that are not exactly face-on to Earth,

satellites on either side of the host galaxy as seen on the sky will in general have opposite velocities relative to the host (that is, the velocities of the satellites will be anti-correlated). This motivates the following simple detection method: for each satellite around a given host, we check whether it possesses a counterpart that is located on the opposite side of the host galaxy to within a certain tolerance angle α (Fig. 1a), and, if it does, we determine whether the pair has correlated or anti-correlated velocities. With circular orbits, no contamination and perfect data, all pairs will be anti-correlated if they all belong to co-rotating planes.

As a control, we first apply this test to the large Millennium II simulation (MS2) of structure formation and evolution^{23,24}, which reflects our best theories of galaxy formation in Λ cold dark matter (Λ CDM) cosmology²⁵. We find that diametrically opposite pairs of bright satellites selected from that simulation (Methods) show no kinematic coherence, with roughly equal numbers of correlated and anti-correlated pairs for all α (filled circles in Fig. 1b). However, with a very different satellite selection strategy, a slight preference for co-rotating satellites can be found²⁶. To analyse the behaviour of our statistic in the presence of a contaminating background, we forced different fractions of satellites around M31-like hosts in MS2 (Methods) to lie within a randomly chosen rotating plane. Figure 1c shows the anti-correlation of the satellite pairs as a function of the dominance of planar configurations; evidently, a measure of the fraction of anti-correlated pairs in real galaxies has the potential to reveal whether planar satellite alignments are common.

We therefore applied this test to the SDSS, which at present gives the most complete view of the nearby universe. Because we wish to investigate the environment around galaxies similar to the Milky Way and Andromeda, we select isolated host galaxies (no brighter neighbour within a distance of 500 kpc or with a velocity that differs by less than $1,500 \text{ km s}^{-1}$) with r -band absolute magnitudes in the range $-23 \text{ mag} \leq M_r \leq -20 \text{ mag}$ from the NYU Value-Added Galaxy Catalog²⁷. Cosmological parameters from the European Space Agency's Planck mission are assumed²⁸. To ensure a clean sample of satellites, we select hosts up to a redshift of $z = 0.05$ (beyond this, few faint satellites are detected), remove hosts closer than $z = 0.002$ to avoid noisy measurements, and remove all galaxies with velocity uncertainties greater than 25 km s^{-1} . The satellites themselves are any galaxies one magnitude or more fainter than the host, but brighter than $M_r = -16 \text{ mag}$, within the radial range $20 \text{ kpc} < R < 150 \text{ kpc}$ (again to be similar to the M31 analysis⁵). We further require that the satellites at projected distance R lie within a velocity of $300 \exp[-(300 \text{ kpc}/R)^{0.8}] \text{ km s}^{-1}$ (Methods and Extended Data Fig. 1).

As for the MS2 analysis, we retain only those satellites whose direction of motion with respect to their hosts is well resolved; because we impose an upper velocity uncertainty of 25 km s^{-1} for both hosts and satellites, we require a minimum velocity difference of $25\sqrt{2} \text{ km s}^{-1}$. There are 380 galaxy systems in the SDSS that pass these requirements.

Various choices for α are examined in Fig. 2. As our toy model shows (Methods and Fig. 1b), the highest contrast between anti-correlated and correlated satellite pairs should be found for small α . There is an inevitable trade-off between the number of satellite pairs that pass the selection criteria and the contamination fraction suffered by the sample.

¹Lycée international des Pontonniers, 1 rue des Pontonniers, F-67000 Strasbourg, France. ²Observatoire astronomique de Strasbourg, Université de Strasbourg, CNRS, UMR 7550, 11 rue de l'Université, F-67000 Strasbourg, France. ³Sydney Institute for Astronomy, School of Physics, A28, The University of Sydney, New South Wales 2006, Australia.

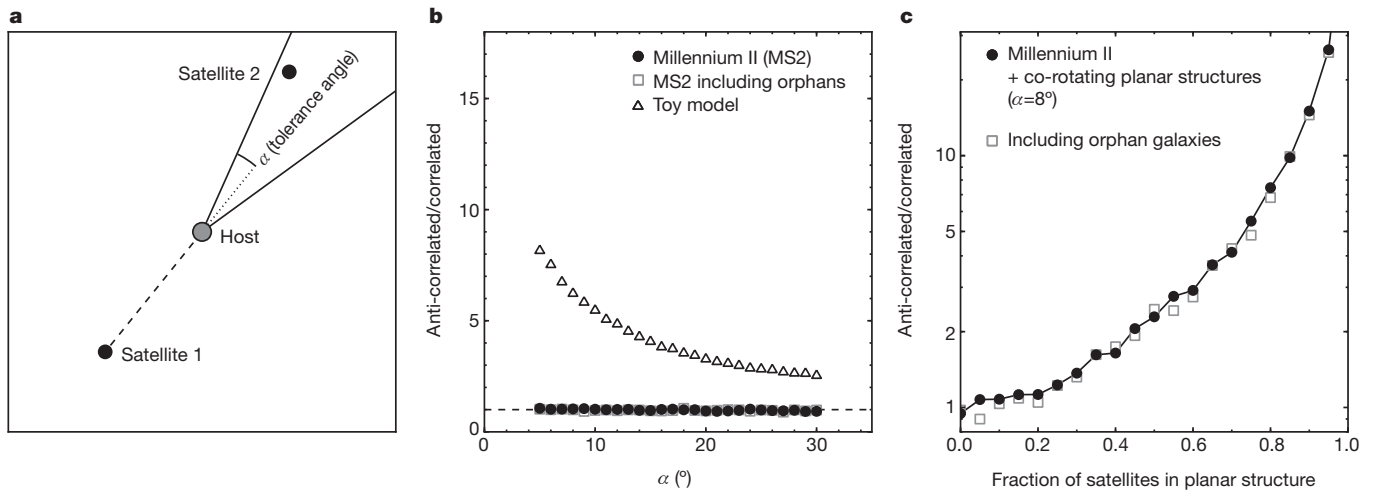


Figure 1 | Satellite correlation test. **a**, Sketch of the satellite selection process. **b**, The fraction of anti-correlated pairs to correlated satellite pairs in MS2 (rejecting or including ‘orphan’ galaxies; see Methods) is consistently very close to 1, independently of α . However, the simple toy model (Methods) shows a decrease in the ratio with increasing α . **c**, Fraction of anti-correlated galaxy pairs

Strict selection gives good contrast, but poor statistics; lenient selection gives poor contrast, but good statistics. Optimal significance will therefore lie at an intermediate tolerance angle, but, given the unknown density and kinematic properties of both the normal and putative disk-like satellite populations, we believe that the best strategy is to allow the data themselves to guide our choice of α . Figure 2a shows that with $\alpha = 8^\circ$, 20 of 22 pairs are anti-correlated, implying a significance of 99.994% ($>4\sigma$); these systems are listed in Table 1. (With a less strict velocity cut-off, of $20\sqrt{2} \text{ km s}^{-1}$, 21 of 23 pairs are anti-correlated.) High significance is found out to $\alpha = 15^\circ$, although, as expected in the presence of non-planar ‘contaminants’, the significance decreases with increasing α . By comparison with the simulations where a disk population was added to MS2 (Fig. 1c), the observed ratio of anti-correlated pairs to correlated pairs (>2.7 at 99% confidence) at $\alpha = 8^\circ$ suggests that $>60\%$ of satellites reside in planes, although we stress that this constraint is weak because this disk model is simplistic. Thus, we have found that the average giant galaxy in the SDSS is consistent with our M31 template. Although the SDSS spectroscopic observing strategy produced certain spatial biases²⁹,

as a function of the fraction of satellites in the co-rotating planar population (using $\alpha = 8^\circ$, the most significant peak in Fig. 2c). In the absence of a planar component, equal numbers of correlated and anti-correlated satellites should be detected. However, the ratio increases as expected as we increase the fraction of satellites in the planar component.

it seems extremely improbable that such biases could artificially cause an overabundance of anti-correlated satellite pairs as found here.

Figure 3 highlights a possible correlation between the direction defined by the satellite alignment and the large-scale structure surrounding the hosts. An elongated overdensity of galaxies appears to be aligned along the axis of the satellite pair, extending out to ~ 10 times the radial distance of the selected pair (Fig. 3a, b). This is consistent with what we see in the Local Group, where the M31 satellite alignment points to within 1° of the Milky Way. Although these filaments of galaxies are much thicker than the planes around host galaxies, it is possible that this reveals the influence of large-scale structure on the dynamics of the smaller satellite system. Furthermore, in MS2 the larger-scale environment around anti-correlated pairs shows no strong preferential direction, and neither does the environment around SDSS correlated pairs (Fig. 3c). Although it remains possible that the large-scale elongation of the galaxy distribution along the direction of the galaxy pairs is an artefact of the SDSS target selection, this seems unlikely given the random orientation of the satellite pairs on the sky.

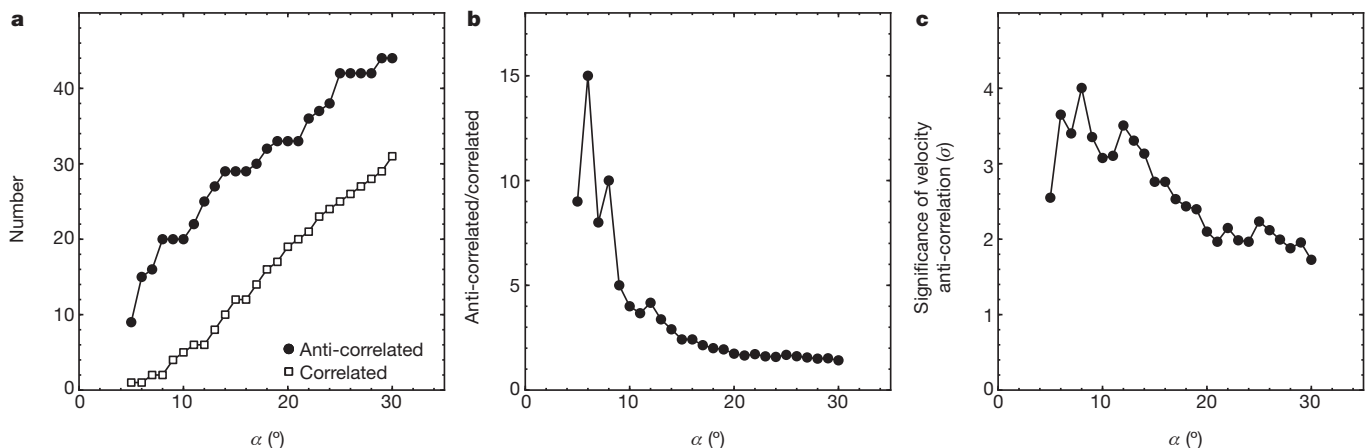


Figure 2 | Anti-correlated satellites in the SDSS. **a**, The number of satellite pairs that have correlated and anti-correlated velocities is shown as function of the tolerance angle. There is a clear surplus of anti-correlated pairs for all angles considered. **b**, The fraction of anti-correlated pairs to correlated satellite pairs shows an overall decrease with increasing tolerance angle, reaching 2.4 at

15° , which we consider the maximum useful opening angle given the low number of satellite pairs in the SDSS. **c**, The significance of the excess of anti-correlated satellite pairs. The most significant peak has significance $>4\sigma$ at an opening angle of 8° .

Table 1 | Host and satellite galaxy parameters

z	RA ^h (°)	dec. ^h (°)	M_r^h (mag)	RA ^{S1} (°)	dec. ^{S1} (°)	M_r^{S1} (mag)	v^{S1} (km s ⁻¹)	RA ^{S2} (°)	dec. ^{S2} (°)	M_r^{S2} (mag)	v^{S2} (km s ⁻¹)	$A L_* $ (10 ¹³ M _⊙ km s ⁻¹ kpc)
0.0324	318.182771	-0.387524	-20.01	318.172302	-0.395010	-17.46	43	318.218964	-0.366205	-18.31	-131	1.2
0.0395	3.140439	-0.048469	-22.00	3.102046	-0.071687	-18.63	-113	3.166931	-0.036225	-19.74	64	5.9
0.0395	3.140439	-0.048469	-22.00	3.102046	-0.071687	-18.63	-113	3.182354	-0.028088	-19.35	66	9.7
0.0197	181.112776	1.895961	-22.84	181.153046	1.892648	-17.90	189	181.054535	1.900154	-19.29	-159	12.1
0.0476	129.700626	4.126124	-22.02	129.688538	4.113726	-20.12	-182	129.715744	4.139352	-19.39	161	8.3
0.0443	230.341251	5.066835	-22.23	230.337646	5.045036	-18.92	107	230.342621	5.082857	-19.08	-116	3.8
0.0227	203.116383	7.316453	-22.46	203.134415	7.294215	-20.49	-52	203.106171	7.331840	-18.68	195	7.3
0.0197	210.702609	9.341378	-22.08	210.619705	9.315668	-17.20	-180	210.779785	9.363044	-20.82	86	31.4
0.0360	244.995300	10.493549	-22.41	244.980530	10.486047	-18.54	-57	245.008408	10.499287	-21.12	231	30.8
0.0266	149.548260	15.946962	-21.78	149.566422	15.919011	-18.66	211	149.534317	15.963491	-20.39	-96	9.8
0.0390	241.663921	17.761186	-22.50	241.681305	17.727509	-18.62	-180	241.651321	17.796453	-20.56	40	14.7
0.0208	176.008968	19.949823	-22.68	176.050766	19.942766	-17.97	-41	175.986832	19.955805	-18.79	62	1.4
0.0371	156.234227	20.402742	-22.05	156.282028	20.396730	-18.59	53	156.221756	20.403419	-19.77	-125	5.8
0.0170	157.320107	26.099230	-21.48	157.346039	26.070440	-18.94	-107	157.256622	26.153496	-18.20	87	2.8
0.0272	168.791680	31.033707	-22.27	168.785461	31.002020	-20.08	212	168.804108	31.079819	-17.95	-166	23.8
0.0428	170.378180	33.957669	-22.57	170.338287	33.955376	-19.55	-111	170.402756	33.958282	-18.79	-98	-6.6
0.0277	167.171535	36.161068	-21.50	167.216934	36.116596	-18.07	-90	167.115128	36.206646	-19.52	94	7.8
0.0316	247.216940	42.812009	-21.74	247.204544	42.805214	-19.21	-116	247.243317	42.828125	-20.59	45	7.0
0.0369	164.010031	44.395565	-21.29	163.993439	44.380817	-18.96	111	164.039291	44.428234	-19.13	159	-8.3
0.0375	128.717100	44.637116	-21.91	128.674500	44.593529	-19.51	134	128.751175	44.671597	-19.02	-156	9.9
0.0245	195.793476	47.393865	-21.55	195.792770	47.381126	-19.08	-97	195.788147	47.447929	-18.75	51	2.1
0.0233	143.529090	67.431139	-21.35	143.660706	67.374802	-18.20	-90	143.446182	67.477814	-18.46	52	2.1

Redshifts (z), positions (right ascension (RA) and declination (dec.)), absolute magnitudes (M_r) and radial velocities (v) of the hosts (superscript 'h') and the satellites (superscripts 'S1' and 'S2'), for the sample selected with a tolerance angle of $\alpha = 8^\circ$. The final column lists the sums of the angular momenta of the stellar components of the two satellites: $|L_*| = |L_*^{S1} + L_*^{S2}|$. The constant A takes the value 1 if the pair have anti-correlated velocities and takes the value -1 if the velocities are correlated.

Table 1 also lists the angular momenta of the satellite pairs, calculated using the projected distances, line-of-sight velocities and estimated stellar masses²⁷ of these galaxies. For comparison, the total angular momentum of our Galaxy in stars is $|L_*| \approx 9 \times 10^{13} M_\odot \text{ km s}^{-1} \text{ kpc}$ (where M_\odot is the solar mass and we approximate the Milky Way as an exponential disk³⁰). Thus, the angular momentum contained in the stellar component of the aligned satellites we have identified (mean of the $\alpha = 8^\circ$ sample: $\langle A|L_*| \rangle = 8.3 \times 10^{13} M_\odot \text{ km s}^{-1} \text{ kpc}$, where A indicates sign as explained in Table 1) is comparable to the angular momentum in a giant galaxy's stellar disk. This suggests that these coherent structures make a substantial contribution to the angular momentum budget on galaxy halo scales ($\sim 100 \text{ kpc}$), although a better understanding of their incidence and physical properties is required to quantify their importance.

Our tests were constructed using MS2 as a control sample to predict what should have been a priori expected in Λ CDM cosmology. Just as this paradigm did not predict the planes observed in the Local Group⁷, it did not a priori predict the velocity correlations presented here. It should

be noted, however, that MS2 contains only dark matter, and future large cosmological simulations that include detailed baryonic physics should be performed to see if the discrepancies can be reduced. Although we are still uncertain whether the pairs of satellites detected here actually form part of kinematically coherent planes, their velocity anti-correlation, alignment with larger-scale structures and high angular momentum are all unexpected properties of the Universe that will require explanation.

METHODS SUMMARY

Our test uses satellites that are diametrically opposite each other around their host to quantify the incidence of rotating planar alignments. The signature of coherent rotation is an enhancement in the number of anti-correlated satellites. Using a small tolerance angle (Fig. 1a) and a minimum velocity difference, samples can be generated with a higher probability of containing edge-on planar structures, if they are present. We first test this method on a simple toy model, to show that it behaves as expected for particular choices of the tolerance angle parameter α (Fig. 1b): the contrast of the planar component is seen to decrease with increasing α , suggesting that small values of α should preferably be used for the tests. To construct a more

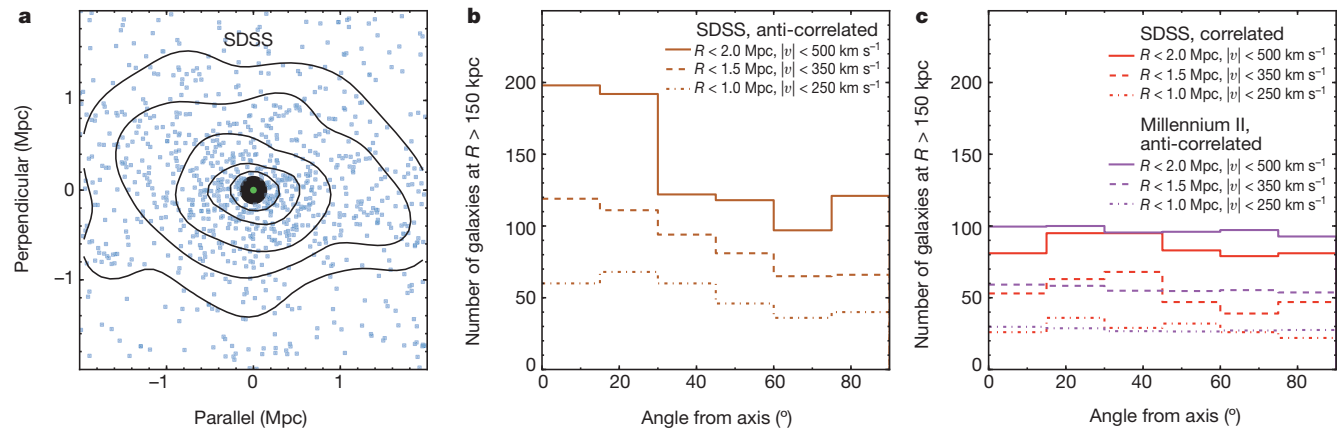


Figure 3 | Correlation with environment. **a**, Superposition of SDSS galaxies (within 500 km s^{-1}) that surround the hosts of the satellite pairs with anti-correlated velocities (using $\alpha = 15^\circ$). Each field is rotated so that the receding satellite lies on the positive abscissa. A clear horizontal feature is found out to $\sim 2 \text{ Mpc}$; this result remains robust for various subsamples and parameter choices. (The black disk shows a radius of 150 kpc .) **b**, The angular distribution

of the galaxies in **a**, rejecting galaxies within $R < 150 \text{ kpc}$. The significances of the peaks for the $R < 1.0, 1.5$ and 2 Mpc samples are $3.7\sigma, 4.8\sigma$ and 7.1σ , respectively. **c**, Applying the same procedure to the region around SDSS correlated pairs (red lines, using $\alpha = 20^\circ$ to build up better statistics) shows minimal correlation, as does the environment around anti-correlated pairs in MS2 (purple lines).

realistic model, we select galaxies and their satellites from the MS2 cosmological simulation, and reassign some of the satellites to planar structures. The selection process for hosts and satellites is kept as close as possible to the process applied to the observed SDSS sample.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 March; accepted 8 May 2014.

Published online 20 July 2014.

1. Kroupa, P., Theis, C. & Boily, C. M. The great disk of Milky-Way satellites and cosmological sub-structures. *Astron. Astrophys.* **431**, 517–521 (2005).
2. Pawlowski, M. S., Pflamm-Altenburg, J. & Kroupa, P. The VPOS: a vast polar structure of satellite galaxies, globular clusters and streams around the Milky Way. *Mon. Not. R. Astron. Soc.* **423**, 1109–1126 (2012).
3. Pawlowski, M. S., Kroupa, P. & Jerjen, H. Dwarf galaxy planes: the discovery of symmetric structures in the Local Group. *Mon. Not. R. Astron. Soc.* **435**, 1928–1957 (2013).
4. Pawlowski, M. S. & Kroupa, P. The rotationally stabilized VPOS and predicted proper motions of the Milky Way satellite galaxies. *Mon. Not. R. Astron. Soc.* **435**, 2116–2131 (2013).
5. Ibata, R. A. *et al.* A vast, thin plane of corotating dwarf galaxies orbiting the Andromeda galaxy. *Nature* **493**, 62–65 (2013).
6. Conn, A. R. *et al.* The three-dimensional structure of the M31 satellite system; strong evidence for an inhomogeneous distribution of satellites. *Astrophys. J.* **766**, 120 (2013).
7. Ibata, R. A. *et al.* A thousand shadows of Andromeda: rotating planes of satellites in the Millennium-II cosmological simulation. *Astrophys. J.* **784**, L6 (2014).
8. Lynden-Bell, D. Dwarf galaxies and globular clusters in high velocity hydrogen streams. *Mon. Not. R. Astron. Soc.* **174**, 695–710 (1976).
9. Adelman-McCarthy, J. K. *et al.* The fourth data release of the Sloan Digital Sky Survey. *Astrophys. J.* **162** (supp.), 38 (2006).
10. Metz, M., Kroupa, P. & Jerjen, H. The spatial distribution of the Milky Way and Andromeda satellite galaxies. *Mon. Not. R. Astron. Soc.* **374**, 1125–1145 (2007).
11. Metz, M., Kroupa, P. & Libeskind, N. I. The orbital poles of Milky Way satellite galaxies: a rotationally supported disk of satellites. *Astrophys. J.* **680**, 287–294 (2008).
12. Metz, M., Kroupa, P. & Jerjen, H. Discs of satellites: the new dwarf spheroidals. *Mon. Not. R. Astron. Soc.* **394**, 2223–2228 (2009).
13. McConnachie, A. W. *et al.* The remnants of galaxy formation from a panoramic survey of the region around M31. *Nature* **461**, 66–69 (2009).
14. Ibata, R. A. *et al.* The large-scale structure of the halo of the Andromeda galaxy. I. Global stellar density, morphology and metallicity properties. *Astrophys. J.* **780**, 128 (2014).
15. Conn, A. R. *et al.* A Bayesian approach to locating the red giant branch tip magnitude. II. Distances to the satellites of M31. *Astrophys. J.* **758**, 11 (2012).
16. Tollerud, E. J. *et al.* The SPLASH survey: spectroscopy of 15 M31 dwarf spheroidal satellite galaxies. *Astrophys. J.* **752**, 45 (2012).
17. Collins, M. L. M. *et al.* A kinematic study of the Andromeda dwarf spheroidal system. *Astrophys. J.* **768**, 172 (2013).
18. Chiboucas, K., Jacobs, B. A., Tully, R. B. & Karachentsev, I. D. Confirmation of faint dwarf galaxies in the M81 group. *Astron. J.* **146**, 126 (2013).
19. Bellazzini, M., Oosterloo, T., Fraternali, F. & Beccari, G. Dwarfs walking in a row. The filamentary nature of the NGC 3109 association. *Astron. Astrophys.* **559**, L11 (2013).
20. Hammer, F. *et al.* The vast thin plane of M31 corotating dwarfs: an additional fossil signature of the M31 merger and of its considerable impact in the whole Local Group. *Mon. Not. R. Astron. Soc.* **431**, 3543–3549 (2013).
21. Walker, M. in *Planets, Stars and Stellar Systems* Vol. 5 (eds Oswalt, T. & Gilmore, G.) 1039–1089 (Springer, 2013).
22. Shaya, E. & Tully, B. The formation of Local Group planes of galaxies. *Mon. Not. R. Astron. Soc.* **436**, 2096–2119 (2013).
23. Boylan-Kolchin, M., Springel, V., White, S. D. M., Jenkins, A. & Lemson, G. Resolving cosmic structure formation with the Millennium-II simulation. *Mon. Not. R. Astron. Soc.* **398**, 1150–1164 (2009).
24. Guo, Q. *et al.* Galaxy formation in WMAP1 and WMAP7 cosmologies. *Mon. Not. R. Astron. Soc.* **428**, 1351–1365 (2013).
25. Komatsu, E. *et al.* Seven-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: cosmological interpretation. *Astrophys. J.* **192** (supp.), 18 (2011).
26. Bahl, H. & Baumgardt, H. A comparison of the distribution of satellite galaxies around Andromeda and the results of Λ CDM simulations. *Mon. Not. R. Astron. Soc.* **438**, 2916–2923 (2014).
27. Blanton, M. R. *et al.* New York University Value-Added Galaxy Catalog: a galaxy catalog based on new public surveys. *Astron. J.* **129**, 2562–2578 (2005).
28. Ade, P. A. R. *et al.* Planck 2013 results. XVI. Cosmological parameters. Preprint at <http://arxiv.org/abs/1303.5076> (2013).
29. Strauss, M. A. *et al.* Spectroscopic target selection in the Sloan Digital Sky Survey: the main galaxy sample. *Astron. J.* **124**, 1810–1824 (2002).
30. McMillan, P. J. Mass models of the Milky Way. *Mon. Not. R. Astron. Soc.* **414**, 2446–2457 (2011).

Acknowledgements Funding for the SDSS and SDSS-II has been provided by the Alfred P. Sloan Foundation, the Participating Institutions, the National Science Foundation, the US Department of Energy, the National Aeronautics and Space Administration, the Japanese Monbukagakusho, the Max Planck Society and the Higher Education Funding Council for England. The Millennium-II simulation databases used in this paper and the web application providing online access to them were constructed as part of the activities of the German Astrophysical Virtual Observatory.

Author Contributions All authors assisted in the development and writing of the paper. N.G.I. primarily contributed to the development of the test for planar alignments, and R.A.I. implemented this test on the SDSS galaxy catalogue.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.G.I. (neil.ibata@gmail.com).

METHODS

Simple kinematic test on diametrically opposite satellites. The simple statistical test we have developed is devised to allow us to quantify the frequency of satellites belonging to disk-like structures. We use primarily the distinctive property of a rotating disk-like structure that objects on opposing sides have anti-correlated velocities. The expectation from observations of M31 is that any such structures are superposed on a 'contaminating' population of 'normal' satellites that appear, to first approximation, to have a spherically symmetric distribution around the host. The presence of such a contaminating population, together with the fact that most galaxies beyond the Local Group have only a small number of satellites with well-measured velocities, means that at present we can test only for the alignments in a statistical manner on a sample of hosts.

Thus, the challenge is to devise a means to enhance the contrast of the putative disk over a potentially dominant spherical population. Because our viewing direction on these distant systems cannot be special, on average, the median inclination of any disk-like structures will be 60° (if we define an edge-on configuration to have inclination 90°). This naturally suggests using a test that makes use of the resulting elongation. However, we can bias a sample towards being more edge-on by selecting those systems with satellites that have radial velocities significantly different from their host galaxy. (Face-on disk-like alignments will have zero velocity difference, as viewed along our line of sight.)

As sketched in Fig. 1a, we consider systems consisting of a massive host galaxy harbouring at least one pair of satellites. Picking each satellite in turn, we determine whether another satellite lies on the opposite side of the host within a tolerance angle α . If both satellites possess a velocity that is significantly different from that of their host, the pair is retained for study. Motivated by the galaxy velocity uncertainties in the SDSS, we selected this minimum velocity difference parameter to be $\Delta v_{\min} = 25\sqrt{2} \text{ km s}^{-1}$ in all calculations presented here (the results are qualitatively very similar for $30 \text{ km s}^{-1} < \Delta v_{\min} < 40 \text{ km s}^{-1}$).

To explore how the method works, we first constructed a very simple test configuration, containing 50% of satellites in a spherical population and 50% in a disk-like structure. Both structures were populated with satellites with uniform probability at radii between 20 and 150 kpc. The satellites in the disk rotate at 40 km s^{-1} , independent of radius, while the spherical population has an isotropic velocity dispersion of 70 km s^{-1} . This toy model is then viewed from a random direction, and two satellites are selected at random beyond a projected radius of 20 kpc. If the two satellites lie on opposite sides of the host to within the chosen tolerance angle, and if both satellites have a velocity difference with respect to their host of more than Δv_{\min} , the pair is retained, and we determine whether the velocities are correlated or anti-correlated. The procedure is repeated 2,000 times for each tolerance angle value.

The open triangles in Fig. 1b show the ratio of the number of anti-correlated to correlated pairs as a function of the tolerance angle α . As this simple model shows, we expect the highest contrast to be found at small tolerance angles. The selection of diametrically opposite galaxies, together with the minimum velocity criterion, ensures that configurations close to edge-on are preferentially selected; for this toy model, the average inclination angle for the $\alpha = 10^\circ$ sample is 80° .

Construction of artificial satellite systems from the Millennium II simulation.

To explore further the reliability of our method to uncover genuine planar satellite alignments, we decided to construct artificial galaxy systems that we could run and test our algorithm on. For this purpose, the Millennium II simulation (in particular with the semi-analytic modelling in ref. 24) provides an ideal view of the expected distribution of galaxies and their satellites in a very large ($10^6 h^{-3} \text{ Mpc}^3$) volume in a ΛCDM universe (h is the dimensionless Hubble parameter). The catalogue lists the absolute r -band magnitudes, total galaxy masses, positions and velocities that are necessary for our comparison with observations.

To create the random views of galaxies derived from the simulation, we proceed as follows. We first choose a random direction from which we will view the galaxies in the Millennium II volume. A list of candidate host galaxies is generated by selecting those objects with absolute magnitudes in the range $-23 \leq M_r \leq -20$ (identical to the selection from the real SDSS data presented in the main text). We examine each of the candidate hosts in turn, placing the host to be studied at $10,000 \text{ km s}^{-1}$ (the mean velocity of the SDSS sample), and then making sure that it appears isolated in projection with no brighter neighbour within 0.5 Mpc, with velocity differences less than $1,500 \text{ km s}^{-1}$.

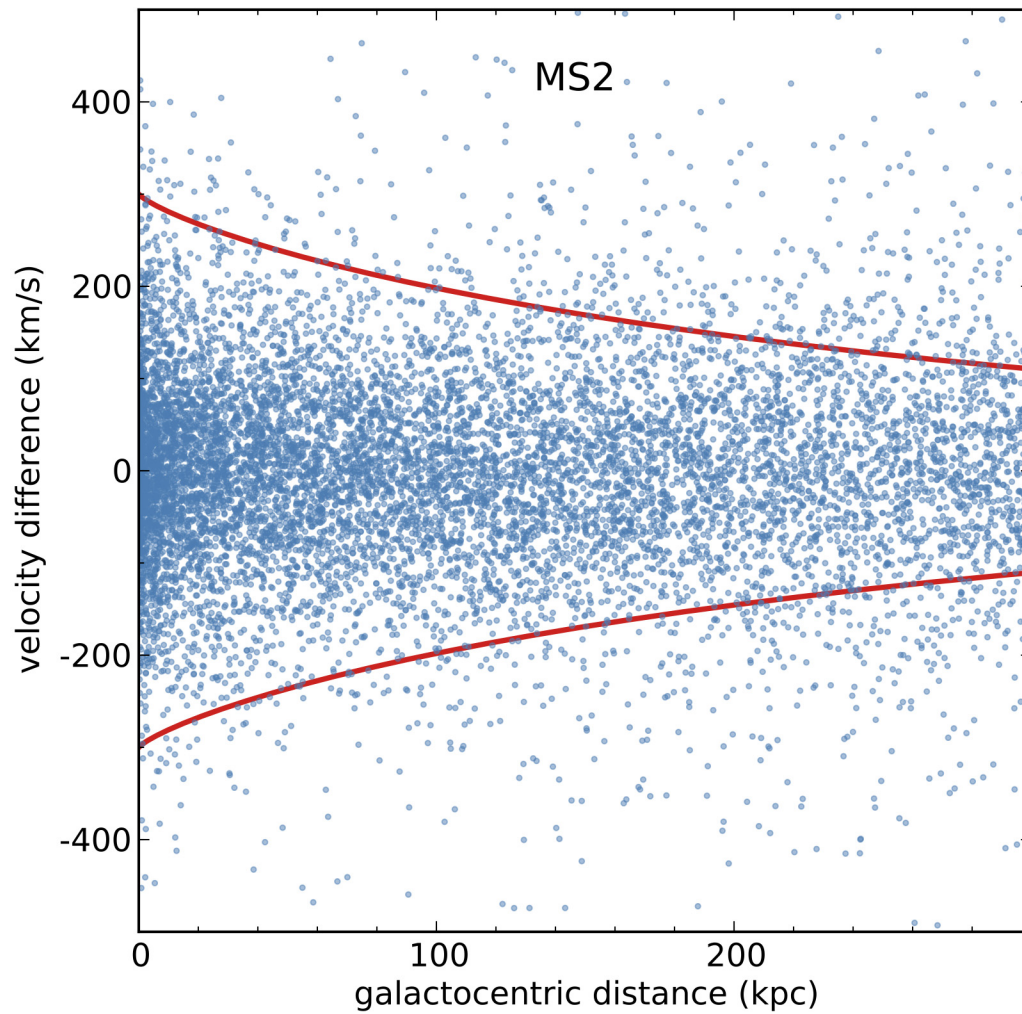
We make a list of all the neighbouring galaxies within a projected distance of 500 kpc and a velocity of $1,500 \text{ km s}^{-1}$ that are at least one magnitude fainter than the host, but that are brighter than $M_r = -16$ mag; we will refer to these objects as 'satellites'. (We reject 'orphan' galaxies, which are systems whose parent subhaloes are no longer resolved in the Millennium II simulation, but, as Fig. 1b, c shows, our results remain qualitatively identical if these objects are included.) For each host, we then randomly draw a vector to define the normal to the planar population, and we go through the list of satellites, randomly assigning them to the planar population, according to the desired planar component fraction that we wish to test for. Clearly, when testing for a planar fraction of zero, the Millennium II positions and velocities remain unaltered. For those satellites that are assigned thus to the planar component, we keep the galactocentric distance that they had in the Millennium II simulation, but place them onto the plane with a random azimuthal angle. The space velocities of the planar satellites are devised to give circular motions in the plane of the alignment, with the total velocity chosen from the circular velocity of a universal halo model³¹ of total mass given by the virial mass of the host.

A Gaussian random velocity of 15 km s^{-1} (a representative value for the SDSS velocity errors) is added to the radial velocity of the host and all its satellites. Having thus reordered the three-dimensional positions of some of the satellites, we filter the sample to keep those objects that lie at projected distances in the range $20 \text{ kpc} < R < 150 \text{ kpc}$ and that have a velocity difference of less than $300 \exp[-(300 \text{ kpc}/R)^{0.8}] \text{ km s}^{-1}$ with respect to the host (Extended Data Fig. 1). The brightest two satellites within the $20 \text{ kpc} < R < 150 \text{ kpc}$ annulus are selected for study. If the two satellites lie on opposite sides of the host to within the chosen tolerance angle, and if both satellites have a velocity difference with respect to their host of more than Δv_{\min} , then the pair is retained and we determine whether the velocities are correlated or anti-correlated.

The entire process is then repeated for all other candidate hosts. We rerun the entire procedure, selecting new initial viewing angles, as many times as necessary until a total of 2,000 satellite pairs have been generated.

An alternative estimate of the fraction of satellites in planes. With the parameter selections detailed in the text, and setting $\Delta v_{\min} = 25\sqrt{2} \text{ km s}^{-1}$, there are 380 galaxy systems in the SDSS. Using $\alpha = 8^\circ$, we find 20 of 380 pairs to have anti-correlated velocities, and 2 of 380 to have correlated velocities; that is, 22 of 380, or 5.8%, of all pairs are found with this tolerance angle. With the unaltered Millennium II simulation (0% in a disk), we find 4.7% of pairs with $\alpha = 8^\circ$; this fraction rises to 4.9% with a 50% disk component, and to 6.7% with a 100% disk component. This suggests that the fraction of satellites in a planar component within the SDSS is greater than 50%, consistent with the estimate given in the text, but the simplicity of the disk model for the planar component prevents us from drawing strong conclusions from this comparison.

31. Navarro, J. F., Frenk, C. S. & White, S. D. M. A universal density profile from hierarchical clustering. *Astrophys. J.* **490**, 493–508 (1997).
32. Li, Y.-S. & White, S. D. M. Masses for the Local Group and the Milky Way. *Mon. Not. R. Astron. Soc.* **384**, 1459–1468 (2008).



Extended Data Figure 1 | Adopted velocity envelope relation. Dots mark the distance–velocity distribution of satellites in the MS2 simulation that surround isolated host galaxies of similar luminosity and mass to the Milky

Way³². The empirical envelope relation shown in red ($300\exp[-(300 \text{ kpc}/R)^{0.8}] \text{ km s}^{-1}$) is used in our analysis as a means to reduce contamination from velocity outliers.

Misaligned protoplanetary disks in a young binary star system

Eric L. N. Jensen¹ & Rachel Akeson²

Many extrasolar planets follow orbits that differ from the nearly coplanar and circular orbits found in our Solar System; their orbits may be eccentric¹ or inclined with respect to the host star's equator^{2,3}, and the population of giant planets orbiting close to their host stars suggests appreciable orbital migration⁴. There is at present no consensus on what produces such orbits. Theoretical explanations often invoke interactions with a binary companion star in an orbit that is inclined relative to the planet's orbital plane^{4,5}. Such mechanisms require significant mutual inclinations between the planetary and binary star orbital planes. The protoplanetary disks in a few young binaries are misaligned^{6–12}, but often the measurements of these misalignments are sensitive only to a small portion of the inner disk, and the three-dimensional misalignment of the bulk of the planet-forming disk mass has hitherto not been determined. Here we report that the protoplanetary disks in the young binary system HK Tauri are misaligned by 60 to 68 degrees, such that one or both of the disks are significantly inclined to the binary orbital plane. Our results demonstrate that the necessary conditions exist for misalignment-driven mechanisms to modify planetary orbits, and that these conditions are present at the time of planet formation, apparently because of the binary formation process.

Although the three-dimensional orbital orientation is not yet measurable for any of the known extrasolar planets, measuring the orientation of protoplanetary disks has the potential to provide information about planetary orbits during the planet formation process. Because these disks are hundreds of astronomical units (1 AU is the average Sun–Earth distance) in diameter, they can be spatially resolved at the 120–160 pc distances of the nearest star-forming regions. If the disks around both stars in a binary system can be shown to be misaligned, then it is clear that both cannot be aligned with the (usually undetermined) binary orbital plane. Indirect evidence of disk misalignment is provided by misaligned jets⁹ and by polarimetry^{13,14}. More directly, images of several young binary systems show that the disk around one star is nearly edge-on to Earth^{6–8,12}. In some of these systems, infrared interferometry or imaging constrains the inclination of the disk around the other star, giving a lower limit on the degree of misalignment of the disks^{8,12}, although the position angle of the disks is uncertain and the direction of rotation is unknown. For systems with detectable millimetre-wavelength emission, measurement of Keplerian rotation in both disks in a binary system provides the opportunity to measure the full three-dimensional orientation of the disks' angular momenta.

One such system is HK Tauri, a young binary system with a projected separation of 2.4 arcsec (ref. 15), which is 386 AU at the distance (161 pc) of this part of the Taurus clouds¹⁶. Age estimates for this system range from 1 to 4 Myr (ref. 17), placing it in the age range at which planet formation is thought to occur. The southern, fainter star, HK Tau B, is surrounded by a disk that blocks the starlight; the disk can thus be clearly seen in scattered-light images at near-infrared and visible wavelengths to be nearly edge-on^{6,7,18}; statistical arguments suggest that the disk is unlikely to be completely aligned with the binary orbit^{6,7}. The northern star, HK Tau A, has strong millimetre-wavelength continuum emission^{19,20},

showing that it too is surrounded by disk material; however, because the disk does not block the starlight, the disk cannot be seen in scattered light owing to the brightness of the star. The striking difference in their visible-light appearance shows that these two disks are not perfectly aligned, but the degree of misalignment has not previously been known because the molecular gas in the northern disk has not been resolved, and a modest inclination difference would be sufficient to explain the different scattered-light morphologies.

We observed HK Tau with the Atacama Large Millimeter Array (ALMA) at frequencies of 230.5 and 345.8 GHz, covering continuum emission from dust and line emission from the carbon monoxide (CO) 2–1 and 3–2 rotational transitions, respectively (Methods). Both the northern and the southern components of the binary are clearly detected in the continuum and the CO line emission. The CO maps (Fig. 1) show the clear signature of rotating disks around each star, with one side of the disk redshifted and the other side blueshifted. The orientations of the two disks are significantly different, with the northern disk axis elongated nearly north–south, roughly 45° from the elongation axis of the southern disk.

We used a Markov chain Monte Carlo analysis to fit disk models to our data to determine the three-dimensional spatial orientations of the disks (Methods). For HK Tau B, the disk orientation is well known from previous scattered-light imaging, and so we adopt from that work¹⁸ an inclination $i = 85^\circ \pm 1^\circ$ and position angle $PA = 42^\circ$. Although the disk inclination and position angle were previously known, our imaging of HK Tau B provides new spatial information because the direction of disk rotation, apparent in Fig. 1b, removes a 180° ambiguity in the disk's orientation. In what follows, we adopt the convention that the position angle is measured east of north and that the quoted position angle is that of the redshifted edge of the disk. Our model fitting reproduces the individual velocity channel images well for both sources (Fig. 2), allowing us to determine the position angle, inclination and direction of rotation of the molecular gas disk in the northern source, HK Tau A. The Markov chain Monte Carlo analysis gives $PA = 352^\circ \pm 3^\circ$ and $i = 43^\circ \pm 5^\circ$ (Extended Data Fig. 1); all uncertainties are given as 68.3% credible intervals.

Measurement of the PA and the inclination of both disks lets us determine the angle between the two disks' angular momentum vectors, with one ambiguity. Equal inclinations on either side of edge-on ($i = 90^\circ$) will appear identical unless it can be determined which edge of the disk is nearer to the observer, for example if high-resolution imaging can determine that one edge of the disk is shadowed by a flared disk edge and the other is not. In the case of HK Tau B, this orientation is known from scattered-light imaging, but it is still unknown for HK Tau A. Combining the observational constraints, we find that the angle between the two disks' angular momentum vectors is $60^\circ \pm 3^\circ$ if both vectors point to the same side of the sky plane, or $68^\circ \pm 3^\circ$ if they do not (Fig. 3).

The clear misalignment between the two disks has important implications for planet migration and orbital evolution, as well as for theories of binary formation. Although nothing in our observations constrains the orientation of the binary orbital plane, the fact that the two disks are misaligned with each other means that they cannot both be aligned

¹Department of Physics and Astronomy, Swarthmore College, 500 College Avenue, Swarthmore, Pennsylvania 19081, USA. ²NASA Exoplanet Science Institute, IPAC/Caltech, Pasadena, California 91125, USA.

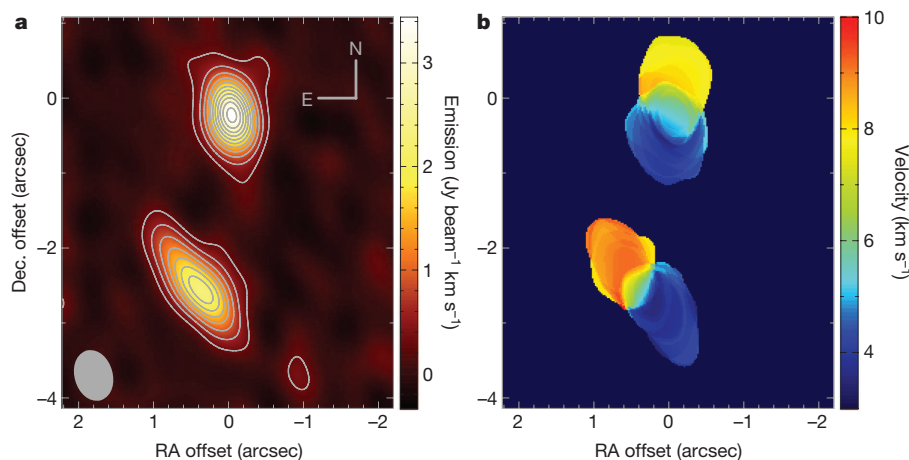


Figure 1 | Observations of the CO(3–2) line in the HK Tau binary system. **a**, Integrated gas emission from each disk, with contours at steps of $0.3 \text{ Jy beam}^{-1} \text{ km s}^{-1}$, three times the root mean squared noise in the maps; the

angular resolution of the observations is shown by the beam size in grey at lower left. RA, right ascension; dec., declination. **b**, Velocity-weighted emission, illustrating the rotation of both disks, and their misaligned orientations.

with the binary orbital plane. At least one of the disks must be misaligned with the binary orbit by 30° (half the total misalignment) or more. The misalignment for one or both disks is probably greater than this, because this minimum misalignment occurs only for one specific orientation of the binary orbit. This misalignment means that planets formed from these disks will be subject to Kozai–Lidov oscillations^{21–23} that may drive changes in their eccentricities and orbital inclinations, or that the disks themselves may be driven into misalignment with the stars' rotation axes⁵. It is sometimes stated that only misalignments greater than the critical angle of 39.2° can cause Kozai–Lidov oscillations^{21,23}, but it has recently been shown that this is not strictly true if the body in the inner orbit is relatively massive or has an eccentric orbit, or both²⁴. In any case, it is quite likely that the inclination relative to the binary orbit exceeds this critical angle for one or both of the disks; only 1.6% of all possible binary orbits are inclined to both disks by less than 39.2° if the disks are misaligned by 60° .

This result is consistent with recent simulations of binary formation^{25–27}, which predict that disks will be misaligned with the binary orbit, especially in systems with orbital semimajor axis lengths greater than 100 AU, where dissipation mechanisms do not act quickly to align the disks with the orbit^{25,28}. In earlier simulations of the formation of individual binary systems from isolated cloud cores, the level of misalignment depended on the choice of initial conditions²⁵. However, more recent simulations^{26,27} focus on the formation of entire clusters and thus do not presuppose specific initial conditions (or even a particular formation mechanism) for an individual binary²⁹. In the cluster simulations of ref. 26, all binary systems with orbital semimajor axes greater than 30 AU have disks that are misaligned with each other, with a mean angle of $70^\circ \pm 8^\circ$. The misalignment we observe here is thus consistent with formation by means of turbulent fragmentation rather than disk instability³⁰.

Although it remains to be seen how the protoplanetary disks in a statistical sample of young binary systems are oriented, it is suggestive

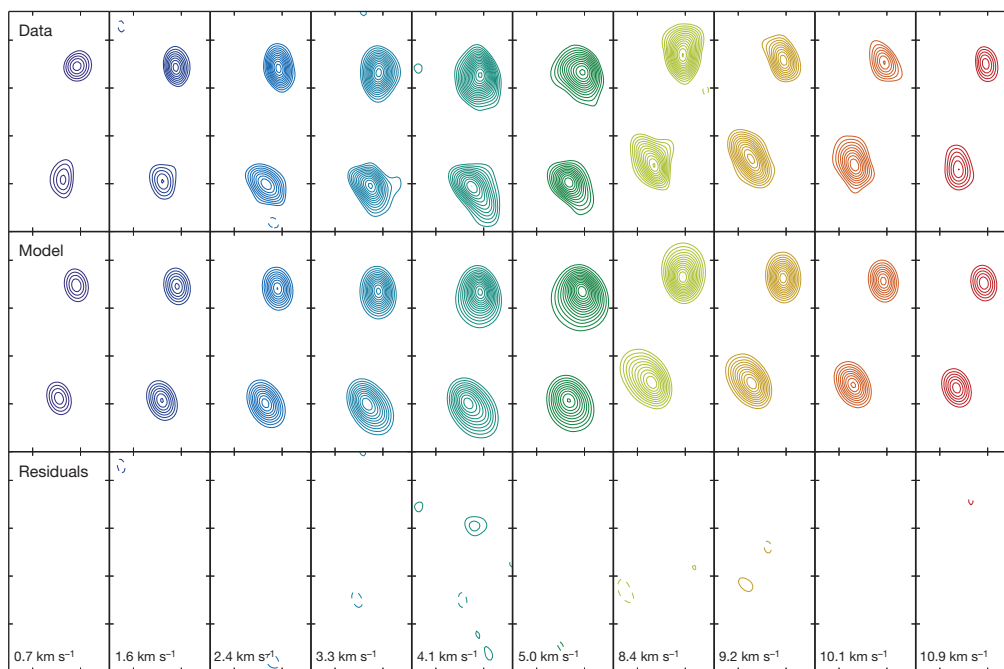


Figure 2 | Data, best-fit model and data–model difference for the disks around HK Tau A and B. Contours are in steps of 28 mJy, the root mean squared noise in the map, starting at three times that value. Negative contours are dashed. North is up and east is to the left, with tickmarks at 1 arcsec

intervals. Three channels near the line centre, which has a velocity of 6.1 km s^{-1} , are omitted from the figure and from calculating the chi-squared statistic in the modelling, owing to absorption from the surrounding cloud.

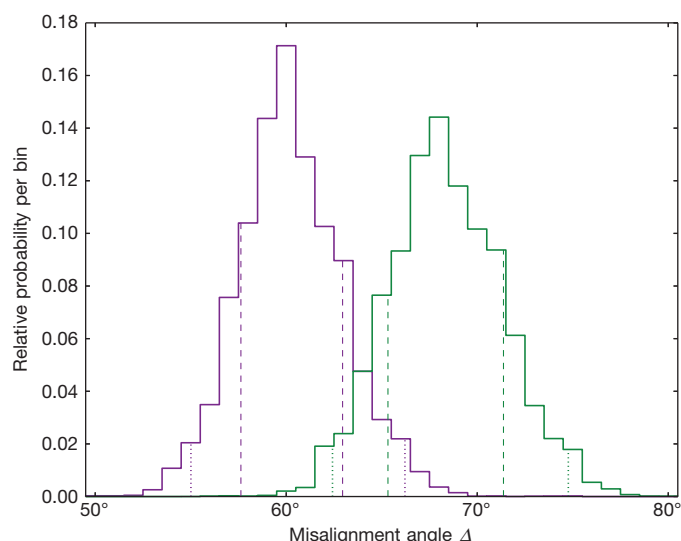


Figure 3 | Posterior probability distribution for the angle Δ between the two disks' angular momentum vectors. The purple histogram is for the case where both disks' vectors are on the same side of the sky plane; the green histogram is for the case where they are on opposite sides of the sky plane. The 68.3% and 95.4% credible intervals are shown by dashed and dotted lines, respectively. See Methods for the definition of Δ .

that in the handful of systems where this measurement has been made, the misalignments are large. If this is a common outcome of the binary formation process, and especially if it extends to lower-mass binary companions (which may easily go undetected), then perturbations by distant companions may account for many of the orbital properties that make the present sample of extrasolar planets so unlike the planets of our own Solar System.

METHODS SUMMARY

The CO(2–1) and CO(3–2) ALMA observations of HK Tau were calibrated using standard techniques. The antenna configuration yielded respective spatial resolutions (here defined by beam sizes from the CLEAN algorithm for image reconstruction) of $1.06 \text{ arcsec} \times 0.73 \text{ arcsec}$ and $0.69 \text{ arcsec} \times 0.51 \text{ arcsec}$ and spectral resolutions of 1.3 km s^{-1} and 0.85 km s^{-1} in the two bands. To determine the disk orientations, we calculated azimuthally symmetric, vertically isothermal parameterized disk models using a Monte Carlo radiation transfer code, and then sampled the model images at the same spatial frequencies and velocities as the observations, allowing us to compare models directly with the calibrated complex visibilities (that is, the Fourier transform of the sky brightness distribution) recorded by the interferometer. A Bayesian Markov chain Monte Carlo analysis yielded posterior probability distributions for the disk parameters.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 November 2013; accepted 20 May 2014.

- Wu, Y. & Murray, N. Planet migration and binary companions: the case of HD 80606b. *Astrophys. J.* **589**, 605–614 (2003).
- Winn, J. N., Fabrycky, D., Albrecht, S. & Johnson, J. A. Hot stars with hot Jupiters have high obliquities. *Astrophys. J.* **718**, L145–L149 (2010).
- Albrecht, S. *et al.* Obliquities of hot Jupiter host stars: evidence for tidal interactions and primordial misalignments. *Astrophys. J.* **757**, 18 (2012).
- Fabrycky, D. & Tremaine, S. Shrinking binary and planetary orbits by Kozai cycles with tidal friction. *Astrophys. J.* **669**, 1298–1315 (2007).
- Batygin, K. A primordial origin for misalignments between stellar spin axes and planetary orbits. *Nature* **491**, 418–420 (2012).

- Stapelfeldt, K. R. *et al.* An edge-on circumstellar disk in the young binary system HK Tauri. *Astrophys. J.* **502**, L65–L69 (1998).
- Koresko, C. D. A circumstellar disk in a pre-main-sequence binary star. *Astrophys. J.* **507**, L145–L148 (1998).
- Roccatagliata, V. *et al.* Multi-wavelength observations of the young binary system Haro 6–10: the case of misaligned discs. *Astron. Astrophys.* **534**, A33 (2011).
- Bohm, K. H. & Solf, J. A sub-arcsecond-scale spectroscopic study of the complex mass outflows in the vicinity of T Tauri. *Astrophys. J.* **430**, 277–290 (1994).
- Duchêne, G., Ghez, A. M., McCabe, C. & Ceccarelli, C. The circumstellar environment of T Tauri S at high spatial and spectral resolution. *Astrophys. J.* **628**, 832–846 (2005).
- Skemer, A. J. *et al.* Evidence for misaligned disks in the T Tauri triple system: 10 μm superresolution with MMTAO and Markov chains. *Astrophys. J.* **676**, 1082–1087 (2008).
- Ratzka, T. *et al.* Spatially resolved mid-infrared observations of the triple system T Tauri. *Astron. Astrophys.* **502**, 623–646 (2009).
- Monin, J. L., Menard, F. & Duchêne, G. Using polarimetry to check rotation alignment in PMS binary stars. Principles of the method and first results. *Astron. Astrophys.* **339**, 113–122 (1998).
- Jensen, E. L. N., Mathieu, R. D., Donar, A. X. & Dullaghan, A. Testing protoplanetary disk alignment in young binaries. *Astrophys. J.* **600**, 789–803 (2004).
- Moneti, A. & Zinnecker, H. Infrared imaging photometry of binary T Tauri stars. *Astron. Astrophys.* **242**, 428–432 (1991).
- Torres, R. M., Loinard, L., Mioduszewski, A. J. & Rodríguez, L. F. VLBA determination of the distance to nearby star-forming regions. III. HP Tau/G2 and the three-dimensional structure of Taurus. *Astrophys. J.* **698**, 242–249 (2009).
- Andrews, S. M., Rosenfeld, K. A., Kraus, A. L. & Wilner, D. J. The mass dependence between protoplanetary disks and their stellar hosts. *Astrophys. J.* **771**, 129 (2013).
- MCCabe, C. *et al.* Spatially resolving the HK Tau B edge-on disk from 1.2 to 4.7 μm : a unique scattered light disk. *Astrophys. J.* **727**, 90 (2011).
- Jensen, E. L. N. & Akeson, R. L. Protoplanetary disk mass distribution in young binaries. *Astrophys. J.* **584**, 875–881 (2003).
- Duchêne, G., Menard, F., Stapelfeldt, K. & Duvert, G. A layered edge-on circumstellar disk around HK Tau B. *Astron. Astrophys.* **400**, 559–565 (2003).
- Kozai, Y. Secular perturbations of asteroids with high inclination and eccentricity. *Astron. J.* **67**, 591–598 (1962).
- Lidov, M. L. The evolution of orbits of artificial satellites of planets under the action of gravitational perturbations of external bodies. *Planet. Space Sci.* **9**, 719–759 (1962).
- Innanen, K. A., Zheng, J. Q., Mikkola, S. & Valtonen, M. J. The Kozai mechanism and the stability of planetary orbits in binary star systems. *Astron. J.* **113**, 1915–1919 (1997).
- Naoz, S., Farr, W. M., Lithwick, Y., Rasio, F. A. & Teyssandier, J. Secular dynamics in hierarchical three-body systems. *Mon. Not. R. Astron. Soc.* **431**, 2155–2171 (2013).
- Bate, M. R. *et al.* Observational implications of precessing protostellar discs and jets. *Mon. Not. R. Astron. Soc.* **317**, 773–781 (2000).
- Bate, M. R. Stellar, brown dwarf and multiple star properties from a radiation hydrodynamical simulation of star cluster formation. *Mon. Not. R. Astron. Soc.* **419**, 3115–3146 (2012).
- Offner, S. S. R., Klein, R. I., McKee, C. F. & Krumholz, M. R. The effects of radiative transfer on low-mass star formation. *Astrophys. J.* **703**, 131–149 (2009).
- Fragner, M. M. & Nelson, R. P. Evolution of warped and twisted accretion discs in close binary systems. *Astron. Astrophys.* **511**, A77 (2010).
- Clarke, C. J. in *Proc. IAU Symp. S240 Vol. 2* (eds Hartkopf, W. I., Guinan, E. F. & Harmanec, P.) 337–346 (International Astronomical Union, 2007).
- Offner, S. S. R., Kratter, K. M., Matzner, C. D., Krumholz, M. R. & Klein, R. I. The formation of low-mass binary star systems via turbulent fragmentation. *Astrophys. J.* **725**, 1485–1494 (2010).

Acknowledgements We thank S. Schnee for help in reducing the ALMA data, S. Myers and R. Indebetow for assistance with ALMA data analysis, L. Prato for sharing data in advance of publication, and M. Hughes, D. Cohen, S. Gaudi, L. Steuerle Schofield and K. Stassun for discussions. This paper makes use of the following ALMA data: ADS/JAO.ALMA#2011.0.00150.S. ALMA is a partnership of the ESO (representing its member states), the NSF (USA) and the NINS (Japan), together with the NRC (Canada) and the NSC and ASIAA (Taiwan), in cooperation with the Republic of Chile. The Joint ALMA Observatory is operated by the ESO, the AUI/NRAO and the NAOJ. The National Radio Astronomy Observatory is a facility of the National Science Foundation operated under cooperative agreement by Associated Universities, Inc.

Author Contributions E.L.N.J. developed the disk modelling code, ran the models and wrote most of the paper. R.A. initiated the project, reduced the data, wrote the text on the observations and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to E.L.N.J. (ejensen1@swarthmore.edu).

METHODS

We observed HK Tau with the Atacama Large Millimeter Array (ALMA) as part of a survey of pre-main-sequence binaries in the Taurus-Auriga star-forming region³¹. Band-6 observations were taken on 17 November 2012 with 27 antennas and band-7 observations on 16 November 2012 with 28 antennas. The correlator was configured with each of the four basebands covering a total bandwidth of 1.875 GHz with a channel spacing of 488 kHz. In band 6, one of the correlator basebands was set to cover the CO(2–1) transition at 230.5 GHz, whereas in band 7, one baseband covered CO(3–2) at 345.8 GHz. We took one observation of HK Tau in each band, bracketed by observations of the gain calibrator J051002+180041, which measures the phase and amplitude response as a function of time. We calibrated the data for each band separately using the CASA software and scripts provided by the NRAO ALMA centre. The system temperature, water vapour phase corrections and flagging were applied using the standard scripts. The amplitude and phase as functions of frequency were calibrated against J0423–013. The absolute flux calibration used Callisto and the 2012 flux models, which resulted in a zero-spacing flux of 8.54 Jy at 230 GHz and 19.45 Jy at 345 GHz.

We generated continuum and CO images using the CLEAN task within CASA, with a robust beam weighting of -1.0 . These settings resulted in a clean beam size of $1.06 \text{ arcsec} \times 0.73 \text{ arcsec}$ in band 6 and $0.69 \text{ arcsec} \times 0.51 \text{ arcsec}$ in band 7. The continuum flux of HK Tau is sufficient to provide a self-calibration reference, and we applied a phase-only self-calibration using HK Tau as the reference. Given the short time on source, we averaged the continuum data to a single point in calculating the self-calibration corrections. The channel spacing, combined with Hanning smoothing in the correlator, provides a spectral resolution of 0.85 km s^{-1} for the CO(3–2) line and 1.3 km s^{-1} for the CO(2–1) line. The continuum emission is not strong enough to substantially affect the individual channels in the CO data and, thus, we did not subtract it.

The maps show clearly detected CO emission centred at an LSR velocity of roughly 6.1 km s^{-1} . Examination of the individual channels of the CO data shows the presence of foreground absorption in the LSR velocity range of roughly $5\text{--}8 \text{ km s}^{-1}$, consistent with the absorption seen in the single-dish ^{13}CO spectrum³².

To quantify the disk properties, in particular the spatial orientation of each disk, we fitted a series of models to the 345 GHz CO(3–2) data. Following many recent authors, we adopt a form for our disk model that is given by a self-similarity solution for circumstellar disks³³, and use the specific parameterization of ref. 34.

Although circumstellar disks in binary systems may be warped owing to interactions with their stellar companions^{28,35–37}, the amount of warping is predicted to be largest for disks with aspect ratios less than 0.05. In contrast, the HK Tau B disk is relatively thick; with its measured scale height of 3.8 AU at a radius of 50 AU (ref. 18), the HK Tau B disk is predicted to have little or no warping. Assuming that the thickness of the HK Tau A disk is similar, warping should be of minimal importance for these disks, and we thus adopt an azimuthally symmetric disk model.

The gas density distribution in the model is azimuthally symmetric, and given by

$$\rho(r, z) = \frac{\Sigma(r)}{\sqrt{2\pi}H_p(r)} \exp\left[-\frac{1}{2}\left(\frac{z}{H_p(r)}\right)^2\right]$$

where z is the vertical height above the disk midplane, and Σ is the surface density distribution, given by

$$\Sigma(r) = \Sigma_c \left(\frac{r}{r_c}\right)^{-\gamma} \exp\left[-\left(\frac{r}{r_c}\right)^{2-\gamma}\right]$$

where Σ_c is a constant such that the surface density at the characteristic radius r_c is Σ_c/e . H_p is the pressure scale height, assumed to be in hydrostatic equilibrium and thus given by

$$H_p(r) = \left(\frac{k_B T(r)}{\mu m_H} \frac{r^3}{GM_*}\right)^{1/2}$$

where T is the temperature, k_B is Boltzmann's constant, μ is the mean molecular weight of the gas, m_H is the mass of a hydrogen atom and M_* is the mass of the star. The disk is assumed to be vertically isothermal, and the radial temperature profile is assumed to satisfy a power law and is normalized at 10 AU :

$$T(r) = T_{10} \left(\frac{r}{10 \text{ AU}}\right)^{-q}$$

Because the ambient radiation in the molecular cloud heats material even far from any star, we adopt a minimum temperature of 10 K ; that is, the power law above applies only out to the radius where $T(r) = 10 \text{ K}$, beyond which the temperature is constant at 10 K .

We assume that the dust and gas have the same temperature at a given radius, that the gas is in local thermodynamic equilibrium, that the gas-to-dust ratio by

mass is 100 and that the number fraction of CO in the gas is 10^{-4} . With these assumptions, there are six free parameters that characterize the disk emission and kinematics in the model: M_{disk} , r_c , T_{10} , M_* , γ and q . In addition, there are the two orientation parameters for the disk: its position angle PA and its inclination i to the line of sight. It is these latter two properties that are of primary interest to us in determining the disks' misalignment; the other six are varied to reproduce the observed emission adequately, but we make no claim that they represent the true disk properties in detail, given the simplicity of the model and degeneracies between the parameters. We fix the position of each component at the coordinates determined from fits to the velocity-integrated (first-moment) maps of the CO emission, and we fix the line centres for both components at 6.1 km s^{-1} .

To find the distributions of parameter values that fit the data, we calculate a set of model disks using the Monte Carlo radiation transfer code RADMC-3D version 0.35 (ref. 38). The standard approach to comparing models to interferometric data is to transform the model images into complex visibilities in the u - v plane (where the visibility is the Fourier transform of the sky brightness distribution and u and v are the coordinates in that plane) so that they can be compared directly with the calibrated data recorded by the interferometer, without the intervening, nonlinear step of creating an image from the interferometric data. In the case of a binary system where both disks have strong emission, this presents an additional complication; although the two disks are cleanly separated in the image plane, their emission overlaps in the u - v plane. Thus, it is necessary to compute models for both disks to compare models to data in the u - v plane. This increases the number of free parameters for each step in the model-data comparison from 8 to 16, complicating the exploration of the parameter space.

To make this problem more tractable, we pursue a modelling strategy that rests on the assumption that the best-fit disk parameters for one star are uncorrelated with those of the other star, allowing us to fit for only 8 parameters at a time. As a preliminary step, we model the two disks in the HK Tau system individually. For each component of the binary, we use RADMC-3D with the model described above to create a single model disk, with images at different velocities across the CO(3–2) line that are separated by the velocity resolution of our observations. We then use the NRAO software CASA to sample the model image with the same u - v coverage as our ALMA observations, and we create a CLEAN image in exactly the same way as we imaged our observations of HK Tau. The resultant model image is compared with a sub-image of our data with the same field of view, velocity channel spacing and pixel scale, and we calculate χ^2 between model and data. Using this image-plane modelling and the MCMC analysis described in more detail below, we find the model parameters that provide the best fits for the A and B disks in the image plane.

With these disk parameter estimates, we then proceed with the more robust u - v plane modelling. To make the exploration of parameter space tractable, we vary parameters for only one disk at a time. In each model run, we hold constant the 8 parameters for one disk at values previously found to give a good fit, and vary only the 8 parameters for the other disk. We combine the two disk model images (one of which is always the same for a given run) into a single image with the disks centred at the known positions of HK Tau A and B. We then sample this model image with the same projected baselines used in the ALMA observations to generate model visibilities that can be compared directly with the data. We bin the data and models into 0.85 km s^{-1} channels, the spectral resolution of the observations, and exclude the three channels near the line centre (LSR velocity range, $5.4\text{--}7.9 \text{ km s}^{-1}$), where there is significant absorption from the cloud. We then calculate χ^2 between the model and data visibilities, with separate terms in the χ^2 sum for the real and imaginary parts of each visibility point. The 10 channels shown in Fig. 2 (spanning LSR velocities $0.3\text{--}5.4$ and $7.9\text{--}11.3 \text{ km s}^{-1}$) are used in calculating χ^2 .

Because multiple combinations of the model parameters can provide almost equally good fits to the data, and because the parameter space is large, we use MCMC to determine the posterior probability distribution of each parameter. As noted above, in each chain we vary only the 8 parameters for one of the disks. We use the Python code emcee³⁹, which implements an affine-invariant ensemble sampler⁴⁰. For most parameters we use a flat prior probability, with the exception of the inclination, where we use a $\sin(i)$ prior probability to account for the fact that randomly distributed inclinations do not have equal probabilities of a given i . We evaluate the posterior probability of each model as $\exp(-\chi^2/2)$ times the prior probability. We ran several separate chains to explore a variety of starting positions for the disk's free parameters, and different fixed parameters for the other disk. In each chain, the ensemble had 30 'walkers' and ran for at least 500 steps. For each chain, we discarded the first 150 steps (4,500 model evaluations) as 'burn-in' so that the results would be independent of the starting positions chosen. Because the results from different chains were consistent with each other, we combined them to produce our final parameter estimates. Not including the burn-in steps, our final results for HK Tau A and HK Tau B are based on 66,000 and 30,000 model evaluations, respectively. As noted above, in the case of HK Tau B, the position

angle and inclination are well known from scattered-light imaging, and so for HK Tau B we adopt the PA and i values found from previous work in the analysis that follows, combined with our new measurements for HK Tau A.

The key quantity we are interested in determining is the angle Δ between the two disks' angular momentum vectors. It is related to the measured position angles and inclinations through spherical trigonometry by

$$\cos(\Delta) = \cos(i_1) \cos(i_2) + \sin(i_1) \sin(i_2) \cos(\text{PA}_1 - \text{PA}_2)$$

With both inclinations specified in the usual range of 0° to 90° , the above equation effectively assumes that both disks have their angular momentum vectors oriented on the same side of the plane of the sky. For the case where the two vectors are on opposite sides of the sky plane, one i above should be replaced with $180^\circ - i$ if i is defined always to be less than 90° . Here we adopt the convention used in specifying the inclination of visual binary orbits⁴¹, where i ranges from 0° to 180° . In this convention, $i < 90^\circ$ corresponds to the case where the disk's orbital motion is in the direction of increasing position angle, or, equivalently, where the disk's angular momentum vector is inclined by an angle $90^\circ - i$ towards the observer relative to the sky plane. Thus, although our adopted convention for position angle (that of the redshifted edge of the disk) is the same as that typically adopted in previous work⁴², our inclination convention differs.

By this convention, the inclination of the HK Tau B disk is 95° (because it is known from scattered-light images that the northern face of the disk is tilted towards Earth), and the best-fit inclination of the HK Tau A disk could be either $43^\circ \pm 5^\circ$ or $137^\circ \pm 5^\circ$. In practice, the two cases do not yield greatly differing values of Δ because HK Tau B is so close to edge-on.

In the near future, it may be possible to distinguish between these two inclinations for HK Tau A. A recently discovered Herbig–Haro object, HH 678, lies 10 arcmin west of HK Tau⁴³. Its position angle of 267° with respect to HK Tau places it on a line that is nearly perpendicular to the HK Tau A disk, suggesting that it may be associated. If so, the sign of the radial velocity of the Herbig–Haro object would break the inclination degeneracy of the HK Tau A disk.

We used fixed values of the orientation of the HK Tau B disk, and the values of PA and i for HK Tau A from our MCMC chains, to find the posterior distribution of Δ for the two disks (Fig. 3). We take the median of the posterior distribution as the most probable value, and we find the values above and below the median that encompass 34.15% of the total probability in each direction to define the 68.3% credible interval (dashed lines); we similarly calculate the 95.4% credible interval (dotted lines). A plot of the posterior distributions of PA and i for HK Tau A (Extended Data Fig. 1) shows that they are uncorrelated, as expected.

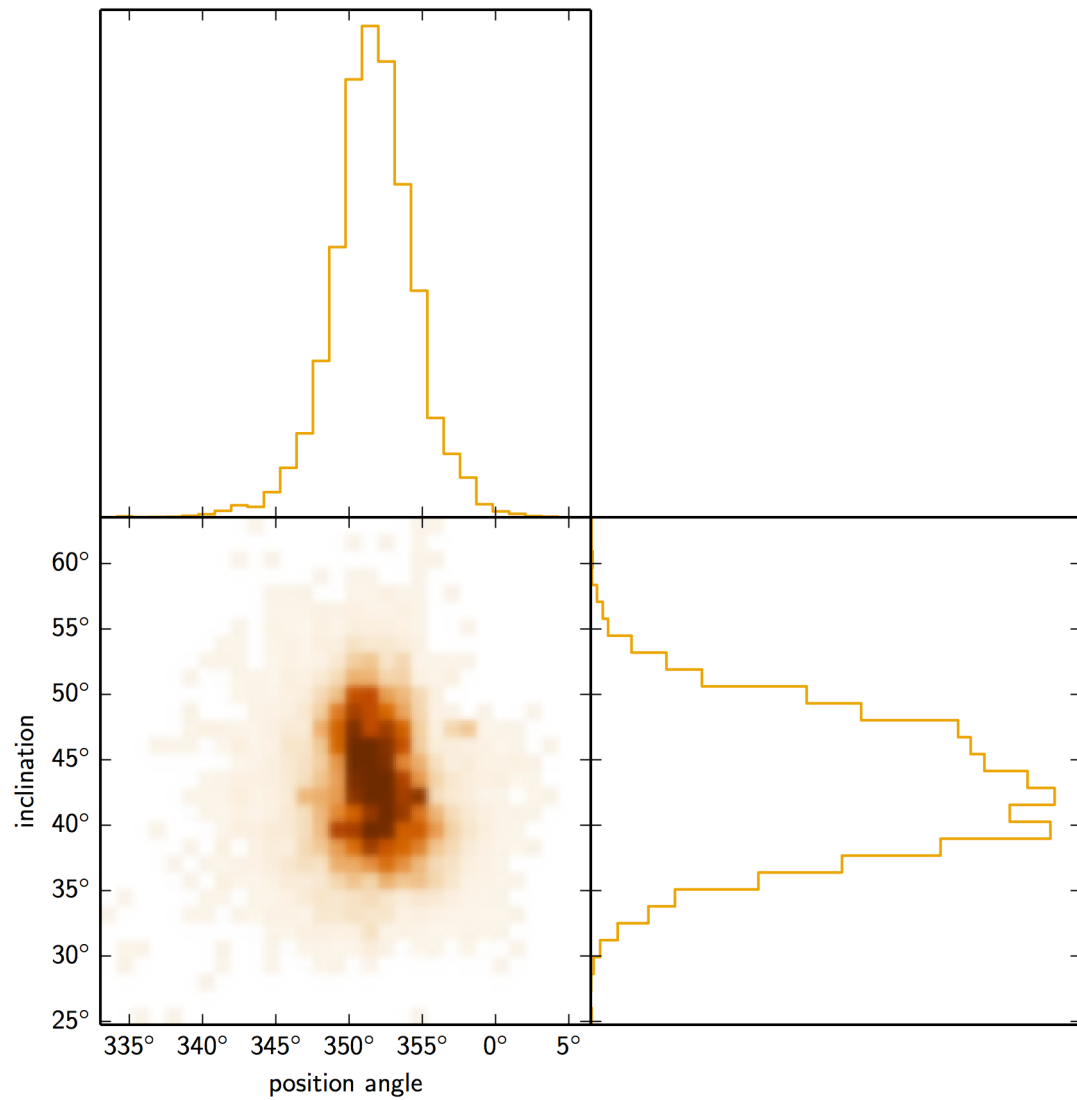
Although our primary focus is the relative orientations of the disks, the modelling used here has the potential to determine other parameters of interest, in particular the stellar mass. Pre-main-sequence stellar mass measurements are of particular interest because they place valuable constraints on pre-main-sequence evolutionary models^{44,45}. Unfortunately, owing to our modest spatial resolution, coupled with the compact size of the HK Tau disks and the cloud absorption over a range of several kilometres per second near the line centre, we are unable to place tight constraints on the stellar masses. Our MCMC analysis yields $M_* = 0.6 \pm 0.1 M_{\text{Sun}}$ for HK Tau A and $M_* = 1.0 \pm 0.1 M_{\text{Sun}}$ for HK Tau B, where the quoted credible intervals do not take into account the uncertainty contribution from the distance to the HK Tau system. The HK Tau A mass is consistent with previous mass estimates from pre-main-sequence evolutionary tracks¹⁷. However, the HK Tau B mass is quite surprising. The published spectral types of HK Tau A and B are M1 and M2, respectively¹³, and near-infrared, high-resolution spectra similarly yield spectral types of M0.5 and M1 for HK Tau A and B (L. Prato, manuscript in preparation). Given its cooler spectral type, and assumed coeval formation, HK Tau B should be less massive than HK Tau A. A possible explanation of the mass discrepancy would

be if HK Tau B were itself a close binary. However, the near-infrared spectra show that the radial velocities of HK Tau A and B are the same to within 1 km s^{-1} , with no evidence of double lines in the spectra of either star (L. Prato, manuscript in preparation).

Thus, we suspect that our stellar mass estimate for HK Tau B may be inaccurate. It may be that our simple models do not adequately reproduce the vertical structure of the disk, which is likely to be much more important in modelling a nearly edge-on disk, like HK Tau B, than in modelling one that is more face-on, like HK Tau A. For example, ALMA science verification data of the disk around HD 163296 show that a vertical temperature gradient is necessary to reproduce the CO emission^{46,47}. It is also possible that the uncertainty in the exact systemic velocity of the system (due to contamination from the molecular cloud) is a factor. Using a fixed systemic velocity parameter may introduce a small bias in the fit parameters, particularly the stellar mass. However, we see no structure in the residuals that would arise from using a systemic velocity far from the correct value.

We emphasize that the position angle and inclination for HK Tau B used in the analysis of disk misalignment were taken from previous scattered-light imaging, and that modelling uncertainties for HK Tau B thus do not affect our main result here. Future ALMA data with better spatial resolution and using an isotopomer that is less sensitive to cloud absorption may help resolve the puzzle of HK Tau B's stellar mass.

31. Akeson, R. L. & Jensen, E. L. N. Circumstellar disks around binary stars in Taurus. *Astrophys. J.* **784**, 62 (2014).
32. Guilloteau, S. et al. A sensitive survey for ^{13}CO , CN, H_2CO , and SO in the disks of T Tauri and Herbig Ae stars. *Astron. Astrophys.* **549**, A92 (2013).
33. Hartmann, L., Calvet, N., Gullbring, E. & D'Alessio, P. Accretion and the evolution of T Tauri disks. *Astrophys. J.* **495**, 385–400 (1998).
34. Rosenfeld, K. A., Andrews, S. M., Wilner, D. J., Kastner, J. H. & McClure, M. K. The structure of the evolved circumbinary disk around V4046 Sgr. *Astrophys. J.* **775**, 136 (2013).
35. Papaloizou, J. C. B. & Terquem, C. On the dynamics of tilted discs around young stars. *Mon. Not. R. Astron. Soc.* **274**, 987–1001 (1995).
36. Larwood, J. D., Nelson, R. P., Papaloizou, J. C. B. & Terquem, C. The tidally induced warping, precession and truncation of accretion discs in binary systems: three-dimensional simulations. *Mon. Not. R. Astron. Soc.* **282**, 597–613 (1996).
37. Lubow, S. H. & Ogilvie, G. I. On the tilting of protostellar disks by resonant tidal effects. *Astrophys. J.* **538**, 326–340 (2000).
38. Dullemond, C. P. RADMC-3D: A multi-purpose radiative transfer tool. *Astrophysics Source Code Library* 1202. 015, <http://asterisk.apod.com/viewtopic.php?f=35&t=27484> (2012).
39. Foreman-Mackey, D., Hogg, D. W., Lang, D. & Goodman, J. emcee: the MCMC hammer. *Publ. Astron. Soc. Pacif.* **125**, 306–312 (2013).
40. Goodman, J. & Weare, J. Ensemble samplers with affine invariance. *Commun. Appl. Math. Comput. Sci.* **5**, 65–80 (2010).
41. Heintz, W. D. *Double Stars* 32 (Reidel, 1978).
42. Piétu, V., Dutrey, A. & Guilloteau, S. Probing the structure of protoplanetary disks: a comparative study of DM Tau, LkCa 15, and MWC 480. *Astron. Astrophys.* **467**, 163–178 (2007).
43. Bally, J., Walawender, J. & Reipurth, B. Deep imaging surveys of star-forming clouds. V. New Herbig–Haro shocks and giant outflows in Taurus. *Astron. J.* **144**, 143 (2012).
44. Simon, M., Dutrey, A. & Guilloteau, S. Dynamical masses of T Tauri stars and calibration of pre-main-sequence evolution. *Astrophys. J.* **545**, 1034–1043 (2000).
45. Rosenfeld, K. A., Andrews, S. M., Wilner, D. J. & Stempels, H. C. A. Disk-based dynamical mass estimate for the young binary V4046 Sgr. *Astrophys. J.* **759**, 119 (2012).
46. Rosenfeld, K. A., Andrews, S. M., Hughes, A. M., Wilner, D. J. & Qi, C. A spatially resolved vertical temperature gradient in the HD 163296 disk. *Astrophys. J.* **774**, 16 (2013).
47. de Gregorio-Monsalvo, I. et al. Unveiling the gas-and-dust disk structure in HD 163296 using ALMA observations. *Astron. Astrophys.* **557**, A133 (2013).



Extended Data Figure 1 | Posterior probability distributions for the position angle and inclination of the disk around HK Tau A.

Mapping the optimal route between two quantum states

S. J. Weber¹, A. Chantasri², J. Dressel³, A. N. Jordan^{2,4}, K. W. Murch⁵ & I. Siddiqi¹

A central feature of quantum mechanics is that a measurement result is intrinsically probabilistic. Consequently, continuously monitoring a quantum system will randomly perturb its natural unitary evolution. The ability to control a quantum system in the presence of these fluctuations is of increasing importance in quantum information processing and finds application in fields ranging from nuclear magnetic resonance¹ to chemical synthesis². A detailed understanding of this stochastic evolution is essential for the development of optimized control methods. Here we reconstruct the individual quantum trajectories^{3–5} of a superconducting circuit that evolves under the competing influences of continuous weak measurement and Rabi drive. By tracking individual trajectories that evolve between any chosen initial and final states, we can deduce the most probable path through quantum state space. These pre- and post-selected quantum trajectories also reveal the optimal detector signal in the form of a smooth, time-continuous function that connects the desired boundary conditions. Our investigation reveals the rich interplay between measurement dynamics, typically associated with wavefunction collapse, and unitary evolution of the quantum state as described by the Schrödinger equation. These results and the underlying theory⁶, based on a principle of least action, reveal the optimal route from initial to final states, and may inform new quantum control methods for state steering and information processing.

Our experiment focuses on the dynamics of two quantum levels of a superconducting circuit (a quantum bit, or qubit), which can be continuously measured and excited by microwave pulses. To access individual quantum trajectories, we make use of the fact that fully projective measurement (or wavefunction collapse) happens over an average time-scale τ controlled by the interaction strength between the system and the detector. By recording the measurement signal with high fidelity in time steps much shorter than τ , we realize a continuous sequence of weak measurements and track the qubit state as it evolves in a single experimental iteration. Individual weak measurements have been recently used in atomic physics experiments that probe wavefunction collapse⁷ and demonstrate state stabilization⁸. In the domain of superconducting circuits, weak measurements⁹ have only recently been realized, owing to the challenge associated with high-fidelity detection of microwave signals near the single-photon level. Advances in superconducting parametric amplifiers have enabled continuous feedback control^{10–12}, the observation of individual quantum trajectories^{13,14}, the determination of weak values^{15,16} and the entanglement of qubits^{17,18}.

In previous work¹³, we demonstrated the ability to track individual quantum trajectories using continuous quantum non-demolition weak measurement. To fully understand the nature of these trajectories, it is necessary to explore their statistical and dynamical properties. Here, by examining a large number of trajectories, we gain insight into the conditional dynamics of open quantum systems. We consider the subset of trajectories that end in a particular final state, which reveals the most probable path connecting two points in quantum state space. Furthermore, whereas previous work¹³ considered only the case of continuous

measurement, we now introduce a concurrent drive at the qubit frequency, resulting in Rabi oscillations that turn qubit state populations into coherences, and vice versa. We are able to track quantum trajectories that exhibit dynamics associated with both measurement backaction and unitary evolution, and find that our theoretical formalism⁶ quantitatively describes the family of trajectories that connect two points in quantum state space.

Our experiment consists of a superconducting transmon circuit¹⁹ dispersively coupled to a waveguide cavity²⁰ (Fig. 1a). Considering only the two lowest levels of the transmon as a qubit, our system is described by the Hamiltonian $H = H_0 + H_{\text{int}} + H_R$, where

$$H_{\text{int}} = -\hbar\chi a^\dagger a \sigma_z$$

$$H_R = \hbar\frac{\Omega}{2}\sigma_y$$

and where H_0 describes the qubit and cavity energy and decay terms. Here \hbar is the reduced Planck's constant, a^\dagger and a are respectively the creation and annihilation operators for the cavity mode, and σ_y and σ_z are qubit Pauli operators. The Hamiltonian H_R describes a microwave drive at the qubit transition frequency, which induces unitary evolution of the qubit state characterized by the Rabi frequency Ω , and H_{int} is the interaction term, characterized by the dispersive coupling rate

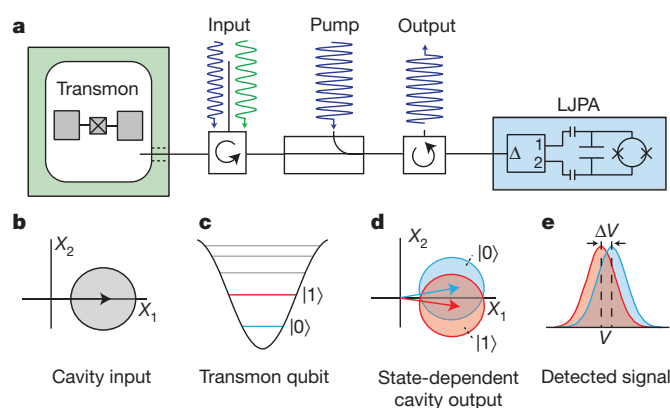


Figure 1 | Set-up. **a**, A transmon circuit is dispersively coupled to a three-dimensional copper waveguide cavity. Microwave signals that reflect off the cavity port are amplified by a lumped-element Josephson parametric amplifier²² (LJPA) operating near the quantum limit. **b**, A microwave tone that probes the cavity near resonance is shown as a phasor in the X_1 – X_2 plane, with zero-point quantum fluctuations shown by the shaded region. **c**, Ground and excited energy levels are shown on the transmon potential. **d**, The reflected microwave tone acquires a qubit-state-dependent phase shift that is smaller than the quantum fluctuations of the measurement signal. After further amplification, the X_2 quadrature of the measurement tone is digitized. **e**, The measurement is calibrated by examining the distributions of measurement signals for the qubit prepared in the $|0\rangle$ (blue) and $|1\rangle$ (red) states.

¹Quantum Nanoelectronics Laboratory, Department of Physics, University of California, Berkeley, California 94720, USA. ²Department of Physics and Astronomy and Center for Coherence and Quantum Optics, University of Rochester, Rochester, New York 14627, USA. ³Department of Electrical Engineering, University of California, Riverside, California 92521, USA. ⁴Institute for Quantum Studies, Chapman University, University Drive, Orange, California 92866, USA. ⁵Department of Physics, Washington University, St Louis, Missouri 63130, USA.

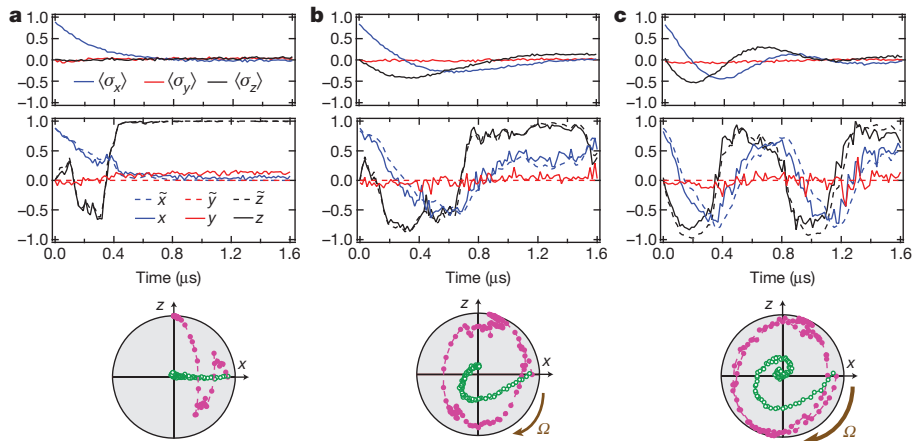


Figure 2 | Quantum trajectories of the quantum state on the Bloch sphere are plotted against time. The upper panels depict the full ensemble evolution. The middle panels depict individual quantum trajectories (dashed curves), with comparison with their tomographic reconstructions (solid curves). At the

bottom, we plot individual trajectories (magenta) and the ensemble averages (green) in the x - z plane of the Bloch sphere. **a–c** correspond to different values of the Rabi drive: $\Omega/2\pi = 0$ MHz (**a**), 0.56 MHz (**b**) and 1.08 MHz (**c**). Here $\tau = 315$ ns and $\Gamma = 3.85 \times 10^6$ s $^{-1}$.

$\chi/2\pi = -0.6$ MHz. This term describes a qubit-state-dependent frequency shift of the cavity, which we use to perform quantum state measurement in our system. We will work in a rotating frame to eliminate the precession of the Bloch vector from the energy level splitting of the qubit. As depicted in Fig. 1b–e, a microwave tone that probes the cavity near its resonance frequency will acquire a qubit-state-dependent phase shift. If the measurement tone is very weak, quantum fluctuations of the electromagnetic mode fundamentally obscure this phase shift, resulting in a partial or weak measurement of the qubit state. We use a near-quantum-limited parametric amplifier^{21,22} to amplify the X_2 quadrature of the reflected signal, which is proportional to the qubit-state-dependent phase shift. After further amplification, we digitize the signal in 16 ns time steps, resulting in a measurement signal $V(t)$. Each time step is small compared with the characteristic measurement time, $\tau = \kappa/16\chi^2\bar{n}\eta_{\text{col}}\eta_{\text{amp}}$, where \bar{n} is the average intracavity photon number, $\kappa/2\pi = 9.0$ MHz is the cavity decay rate and $\eta_{\text{col}}\eta_{\text{amp}}$ is the measurement quantum efficiency²³, which decomposes into separate collection and amplification efficiencies. The characteristic measurement time τ is calibrated by examining (Gaussian) histograms of the measurement results for the qubit prepared in the σ_z eigenstates $|0\rangle$ and $|1\rangle$, and is defined by the time it takes to separate the two distributions by two standard deviations²⁴, $\Delta V = 2\sigma$.

In our experiment, we prepare the qubit in the positive eigenstate of the σ_x Pauli operator (along the x axis of the Bloch sphere), by first making a projective measurement along the z axis and then a $\pi/2$ -rotation about the y axis¹³. By considering only the instances where the measurement result is found to be $|0\rangle$, we herald the preparation of a high-fidelity ground state. Then a measurement tone at 6.8316 GHz continuously probes the cavity for a variable time t , which weakly measures the qubit in the σ_z basis. Finally, we apply further rotations and perform a projective measurement to conduct quantum state tomography. In Fig. 2 (top panels), we show the ensemble-averaged tomography for three different Rabi drive strengths. From these curves, we extract $\Omega/2\pi = 0$, 0.56 and 1.08 MHz and the ensemble coherence decay rate Γ by comparison with theory as discussed in Methods. From Γ , we calculate a total quantum efficiency $\eta_{\text{tot}} = 1/2\tau\Gamma = \eta_{\text{col}}\eta_{\text{amp}}\eta_{\text{env}} = 0.4$, where the last factor indicates the (nearly negligible) extra environmental dephasing $\eta_{\text{env}} = (1 + \kappa/8\chi^2\bar{n}T_2^*)^{-1}$, with $T_2^* = 15$ μ s.

In each iteration of the experiment, we can use the recorded measurement signal to calculate the best estimate for the qubit state conditioned on the measurement record. As discussed in Methods, at each time step we apply a two-step update procedure to track the evolution of the system density matrix ρ . We account for the measurement result using a quantum generalization of Bayes' rule^{23,24}, and we account for the

Rabi drive by applying a unitary rotation. Our finite detector efficiency reflects our imperfect knowledge about the state of the system and results in a decay of coherence given by rate $\gamma = \Gamma - 1/2\tau$. From the density matrix ρ , we calculate expectation values of the Pauli operators conditioned on the measurement signal: $x \equiv \text{tr}[\rho\sigma_x]$, $y \equiv \text{tr}[\rho\sigma_y]$ and $z \equiv \text{tr}[\rho\sigma_z]$ (the components of the Bloch vector).

In Fig. 2a, we display a sample trajectory with no drive ($\Omega = 0$) that shows the stochastic motion of the qubit state as it evolves under measurement and is ultimately projected into the $|0\rangle$ state. As described in Methods, we use conditioned quantum state tomography to reconstruct

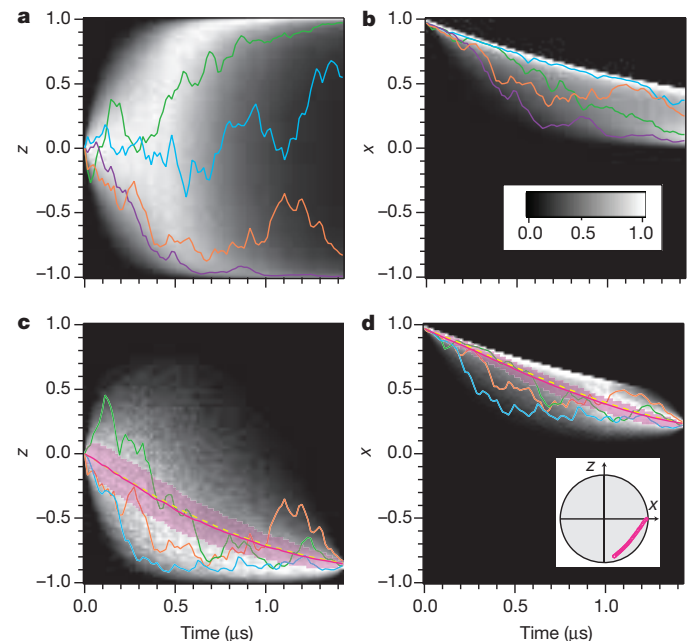


Figure 3 | Greyscale histograms of quantum trajectories in the undriven case. Measurement duration, 1.424 μ s. **a, b**, Histograms of all measured z (**a**) and x (**b**) trajectories, beginning from state ($x_1 = 0.97, z_1 = 0$). Representative trajectories are shown in colour. **c, d**, Histograms of trajectories z (**c**) and x (**d**), conforming to the final chosen boundary condition, $z_F = -0.85 \pm 0.03$. The most likely trajectories from the experimental data are shown as magenta curves, with their standard deviations shown by the magenta bands. The most likely paths in z and x predicted from the theory are shown as yellow dashed curves. Other representative trajectories are shown in other colours. Inset in **d**, the most likely trajectory is plotted on the x - z plane of the Bloch sphere. Here $\tau = 1.25$ μ s and $\Gamma = 0.94 \times 10^6$ s $^{-1}$.

the trajectory. Figure 2b, c demonstrates that we can track the state faithfully in the presence of unitary state evolution induced by a drive at the qubit frequency. The drive induces Rabi oscillations while the system is being continuously measured. The resulting dynamics is fully quantum, going beyond the pure measurement case¹³. These trajectories highlight the stark difference between ensemble dynamics and the dynamics of individual quantum trajectories; whereas the ensemble average decays rapidly to a mixed state, the individual trajectories remain remarkably pure despite the modest quantum efficiency, $\eta_{\text{tot}} = 0.4$.

Using this ability to track individual trajectories starting from a given initial state, we now consider the sub-ensemble of trajectories that arrive at a particular final state at a given time. This sub-ensemble allows us to examine the conditional quantum dynamics of the state that satisfy two boundary conditions, one in the past ('pre-selection') and one in the future ('post-selection'). This is similar to an analysis that leads to 'weak values'^{25–27}, and time-continuous generalizations^{16,28} that consider an additional projective post-selection measurement. In contrast to that approach, we use only a solitary continuous measurement: the pre-selection is just the initial state, and the post-selection is simply whatever the state is when the detector stops measuring. The resulting average of the measurement output gives a 'weak function' that connects the boundary conditions.

To investigate the full ensemble and post-selected sub-ensemble dynamics, we perform 10^5 iterations of the experiment with a measurement duration of $1.424 \mu\text{s}$. For each experiment, we construct the quantum state trajectory by finding x and z for every time step. Figure 3 displays the measurement dynamics for $\Omega = 0$. We consider the sub-ensemble

of trajectories that have final values $(z(1.424 \mu\text{s}), x(1.424 \mu\text{s}))$ within 0.03 of $(x_F, z_F) = (-0.85, 0.23)$. This analysis allows us to examine properties of the conditional trajectories such as the most likely path that connects pre- and post-selected states.

The most likely paths can be theoretically calculated using a stochastic path integral representation of the joint probability of the measurement outcomes at every point in time with boundary condition constraints. The conditional detector backaction on the quantum state can be imposed at every time step with Lagrange multipliers (p_x, p_z) as auxiliary dynamical parameters. Finding the extremum of the stochastic action leads to equations of motion for the optimal path connecting the boundary conditions. As we discuss in Methods, this corresponds to optimizing the total path probability between the states. Because the experiment operates in the x - z plane of the Bloch sphere, the (deterministic) equations of motion for the optimized path are

$$\dot{x} = -\gamma x + \Omega z - xzr/\tau \quad (1)$$

$$\dot{z} = -\Omega x + (1 - z^2)r/\tau \quad (2)$$

$$\dot{p}_x = +\gamma p_x + \Omega p_z + p_x z r/\tau \quad (3)$$

$$\dot{p}_z = -\Omega p_x + (p_x x + 2p_z z - 1)r/\tau \quad (4)$$

where x, z, p_x, p_z and r are now functions of time (with a dot denoting a time derivative) and $r = z + p_z(1 - z^2) - p_x x z$. Here r is the optimal readout and relates to the optimal detector signal as follows: $V_{\text{opt}} = \Delta V r/2$. This rescaling makes r an estimation of z without post-selection (that

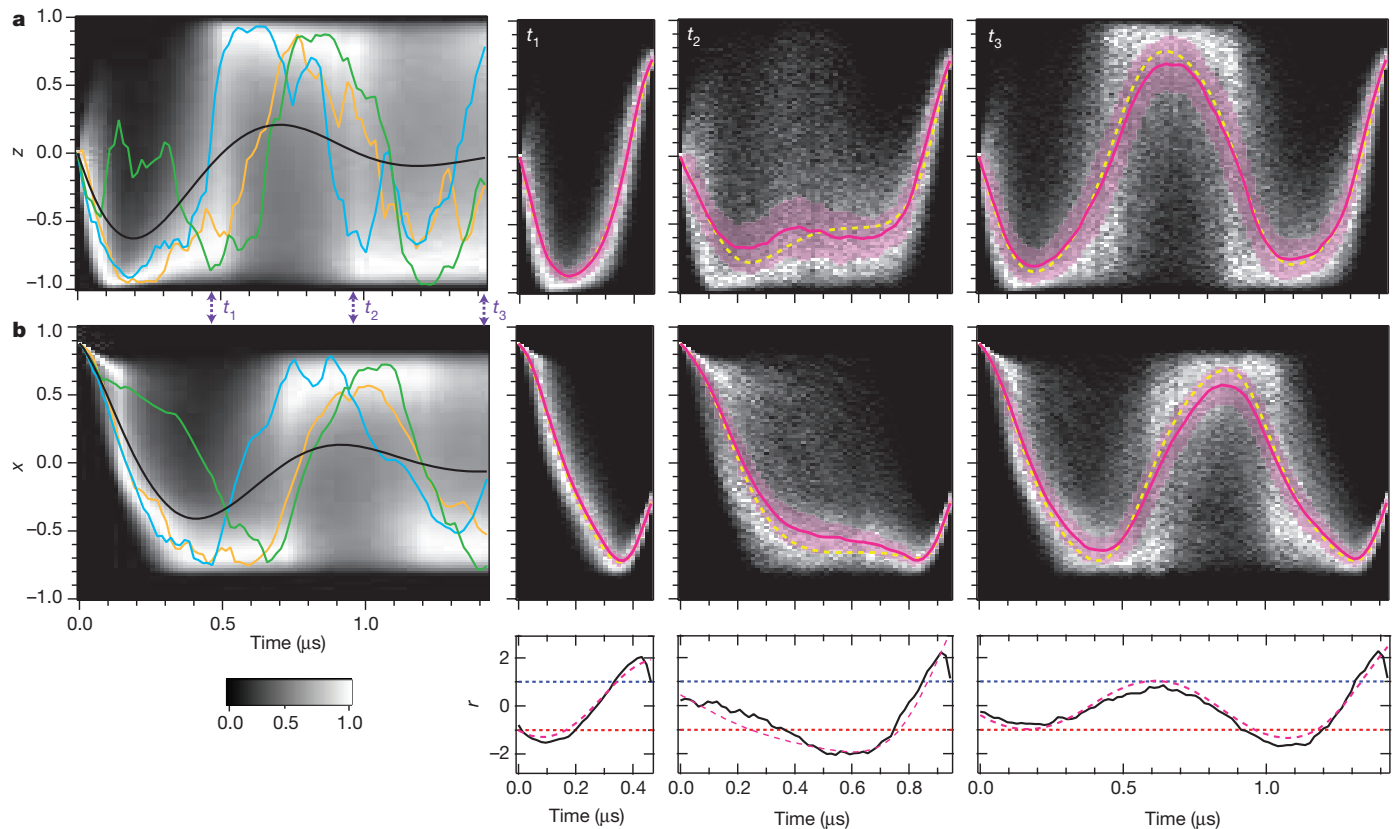


Figure 4 | Greyscale histograms of quantum trajectories in the driven case. The measurements begin at state $(x_1 = 0.88, z_1 = 0)$. Here $\tau = 315 \text{ ns}$, $\Gamma = 3.85 \times 10^6 \text{ s}^{-1}$, $\Omega/2\pi = 1.08 \text{ MHz}$. **a, b**, Histograms for z (**a**) and x (**b**) with representative trajectories plotted in colour and with the average trajectory shown in black. In the other panels in **a** and **b**, we post-select on the final state ($z_F = 0.7, x_F = -0.29$), with a post-selection window of ± 0.08 . Solid magenta curves are the most likely trajectories for the experimental data, and the yellow dashed curves are from the theory. The standard deviations of the

experimentally determined most likely paths are shown by magenta bands. As the time duration between the boundary conditions is increased from $t_1 = 0.464 \mu\text{s}$ to $t_2 = 0.944 \mu\text{s}$ and then to $t_3 = 1.424 \mu\text{s}$, the most likely trajectory connecting the initial and final states changes drastically but is well described by the theory (dashed line). The bottom panels compare the optimal detector signals (r ; dashed lines) with the conditioned average signal (weak functions; black lines).

is, $p_x = p_z = 0$). The solution to these nonlinear equations admits four constants of motion, which permits the imposition of both initial (x_i, z_i) and final (x_f, z_f) boundary conditions.

The equations have a simple analytic solution (\bar{x}, \bar{z}) for $\Omega = 0$. We consider measurement for a time T , starting in the initial state ($x_i = 1, z_i = 0$) and ending in a state (x_f, z_f) (in this particular case, x_f is determined by the choice of z_f). The solution of equations (1)–(4) is ($\bar{x}(t) = e^{-\gamma t} \text{sech}(\bar{r}t/\tau)$, $\bar{z}(t) = \tanh(\bar{r}t/\tau)$), where $\bar{r} = (\tau/T) \tanh^{-1}(z_f)$ is the detector output of maximum likelihood. These solutions are plotted in Fig. 3, showing agreement with the experimentally obtained most likely path (Methods). The most likely times between different boundary conditions are shown in Methods and Extended Data.

In Fig. 4, we display the full ensembles and post-selected ensembles for the driven case ($\Omega/2\pi = 1.08$ MHz). Depending on the amount of time between the initial and final states, the competition between measurement and Schrödinger dynamics produces different (and non-trivial) optimal routes, alternatively showing diffusive Rabi oscillation dynamics and quantum jump dynamics^{6,24,29,30} (where the system is effectively pinned in one of the eigenstates). We compare the experimentally determined most likely trajectories (Methods) with the most likely paths obtained from solving equations (1)–(4). The equations were numerically solved with a shooting method to satisfy both initial and final boundary conditions at different times. These numerical solutions show reasonable agreement with the experimentally determined most likely curves.

In addition to the quantum paths, the solution of equations (1)–(4) also gives the optimal detector response for moving the quantum system to the target state after a given time. We compare these optimal signals to the conditioned average detector signals (weak functions) in Fig. 4. The post-selection allows the conditioned average detector signal r to exceed the usual range of $[-1, 1]$ for z . This behaviour is analogous to that of weak values, which can also lie outside their eigenvalue range²⁷.

The ability to find and verify the most likely path between chosen initial and final quantum states under continuous measurement extends our fundamental understanding of quantum measurement and advances the field of quantum control of individual systems. Our results give deep insight into the quantum dynamics and associated measurement read-out, as revealed by the ability to condition on the final quantum state in the presence of a continuous coherent Rabi drive. The data presented here are in good agreement with our stochastic path integral formalism, predicting the global most likely path, and open the way to solving related optimization problems important to controlling a quantum system. Examples of future applications of this approach, in the specific area of superconducting qubits, include using the continuous measurement results for improved state preparation, state estimation and Hamiltonian parameter estimation. Multiple-qubit architectures can also be fabricated, with each qubit having its own measurement device, enabling optimal continuous control protocols for an ensemble of superconducting qubits with real-time state monitoring. Furthermore, the present work can be extended to solve more general optimization problems in quantum mechanics, such as finding the most likely path from a separable state to a desired entangled state.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 March; accepted 30 May 2014.

1. Vandersypen, L. M. K. & Chuang, I. L. NMR techniques for quantum control and computation. *Rev. Mod. Phys.* **76**, 1037–1069 (2005).
2. Shapiro, M. & Brumer, P. *Quantum Control of Molecular Processes* (Wiley, 2012).
3. Carmichael, H. *An Open Systems Approach to Quantum Optics* (Springer, 1993).

4. Carmichael, H. J., Kochan, P. & Tian, L. in *Proc. Int. Symp. Coherent States: Past, Present, and Future* (eds Feng, D. H., Strayer, M. R. & Klauder, J. R.) 75–91 (World Scientific, 1994).
5. Wiseman, H. & Milburn, G. *Quantum Measurement and Control* (Cambridge Univ. Press, 2010).
6. Chantasri, A., Dressel, J. & Jordan, A. N. Action principle for continuous quantum measurement. *Phys. Rev. A* **88**, 042110 (2013).
7. Guerlin, C. *et al.* Progressive field-state collapse and quantum non-demolition photon counting. *Nature* **448**, 889–893 (2007).
8. Sayrin, C. *et al.* Real-time quantum feedback prepares and stabilizes photon number states. *Nature* **477**, 73–77 (2011).
9. Hatridge, M. *et al.* Quantum back-action of an individual variable-strength measurement. *Science* **339**, 178–181 (2013).
10. Vijay, R. *et al.* Stabilizing Rabi oscillations in a superconducting qubit using quantum feedback. *Nature* **490**, 77–80 (2012).
11. Blok, M. *et al.* Manipulating a qubit through the backaction of sequential partial measurements and real-time feedback. *Nature Phys.* **10**, 189–193 (2014).
12. de Lange, G. *et al.* Reversing quantum trajectories with analog feedback. *Phys. Rev. Lett.* **112**, 080501 (2014).
13. Murch, K. W., Weber, S. J., Macklin, C. & Siddiqi, I. Observing single quantum trajectories of a superconducting qubit. *Nature* **502**, 211–214 (2013).
14. Jordan, A. N. Watching the wavefunction collapse. *Nature* **502**, 177–178 (2013).
15. Groen, J. P. *et al.* Partial-measurement backaction and nonclassical weak values in a superconducting circuit. *Phys. Rev. Lett.* **111**, 090506 (2013).
16. Campagne-Ibarcq, P. *et al.* Observing interferences between past and future quantum states in resonance fluorescence. *Phys. Rev. Lett.* **112**, 180402 (2014).
17. Risté, D. *et al.* Deterministic entanglement of superconducting qubits by parity measurement and feedback. *Nature* **502**, 350–354 (2013).
18. Roch, N. *et al.* Observation of measurement-induced entanglement and quantum trajectories of remote superconducting qubits. *Phys. Rev. Lett.* **112**, 170501 (2014).
19. Koch, J. *et al.* Charge-insensitive qubit design derived from the Cooper pair box. *Phys. Rev. A* **76**, 042319 (2007).
20. Paik, H. *et al.* Observation of high coherence in Josephson junction qubits measured in a three-dimensional circuit QED architecture. *Phys. Rev. Lett.* **107**, 240501 (2011).
21. Castellanos-Beltrán, M. A., Irwin, K. D., Hilton, G. C., Vale, L. R. & Lehnert, K. W. Amplification and squeezing of quantum noise with a tunable Josephson metamaterial. *Nature Phys.* **4**, 929–931 (2008).
22. Hatridge, M., Vijay, R., Slichter, D. H., Clarke, J. & Siddiqi, I. Dispersive magnetometry with a quantum limited SQUID parametric amplifier. *Phys. Rev. B* **83**, 134501 (2011).
23. Korotkov, A. N. Quantum Bayesian approach to circuit QED measurement. Preprint at <http://arxiv.org/abs/1111.4016> (2011).
24. Korotkov, A. N. Continuous quantum measurement of a double dot. *Phys. Rev. B* **60**, 5737–5742 (1999).
25. Watanabe, S. Symmetry of physical laws. Part III. prediction and retrodiction. *Rev. Mod. Phys.* **27**, 179–186 (1955).
26. Aharonov, Y., Bergmann, P. G. & Lebowitz, J. L. Time symmetry in the quantum process of measurement. *Phys. Rev.* **134**, B1410–B1416 (1964).
27. Aharonov, Y., Albert, D. Z. & Vaidman, L. How the result of a measurement of a component of the spin of a spin-1/2 particle can turn out to be 100. *Phys. Rev. Lett.* **60**, 1351–1354 (1988).
28. Williams, N. S. & Jordan, A. N. Weak values and the Leggett-Garg inequality in solid-state qubits. *Phys. Rev. Lett.* **100**, 026804 (2008).
29. Vijay, R., Slichter, D. H. & Siddiqi, I. Observation of quantum jumps in a superconducting artificial atom. *Phys. Rev. Lett.* **106**, 110502 (2011).
30. Gambetta, J. *et al.* Quantum trajectory approach to circuit QED: quantum jumps and the Zeno effect. *Phys. Rev. A* **77**, 012112 (2008).

Acknowledgements We thank A. N. Korotkov, S. G. Rajeev, N. Roch and D. Toyli for discussions. This research was supported in part by the Army Research Office, Office of Naval Research and the Office of the Director of National Intelligence (ODNI), Intelligence Advanced Research Projects Activity (IARPA), through the Army Research Office. All statements of fact, opinion or conclusions contained herein are those of the authors and should not be construed as representing the official views or policies of IARPA, the ODNI or the US government. A.N.J. acknowledges support from NSF grant no. DMR-0844899 (CAREER).

Author Contributions S.J.W. and K.W.M. performed the experiment and analysed the experimental data. J.D. and A.C. wrote the trajectory simulation code. A.C., J.D. and A.N.J. contributed the theory. All work was carried out under the supervision of I.S. All authors contributed to writing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.W.M. (murch@physics.wustl.edu).

METHODS

Device parameters. The qubit consists of two aluminium paddles connected by a double-angle-evaporated aluminium SQUID deposited on double-side-polished silicon. The qubit is characterized by a charging energy $E_C/h = 200$ MHz and a Josephson energy $E_J/h = 11$ GHz. The qubit is operated with negligible flux threading the SQUID loop with a transition frequency $\omega_q/2\pi = 4.01057$ GHz. The qubit is located off centre of a 6.8316 GHz copper waveguide cavity. With the measurement tone on, the qubit transition frequency was a.c.-Stark shifted to 4.00748 GHz. Qubit pulses and drive are performed at the a.c.-Stark-shifted frequency.

The lumped-element Josephson parametric amplifier (LJPA) consists of a two-junction SQUID, formed from 2 μ A Josephson junctions shunted by 3 pF of capacitance, and is flux biased to provide 20 dB of gain at the cavity resonance frequency. The LJPA is pumped by two sidebands equally spaced 300 MHz above and below the cavity resonance.

Experimental set-up. Extended Data Fig. 1 displays a schematic of the experimental set-up. Experimental sequences start with an 800 ns readout to herald the $|0\rangle$ state ($z = +1$), followed by a 16 ns $\pi/2$ -rotation about the y axis to prepare the qubit along the x axis. The state preparation fidelity is 88% for the data shown in Figs 2 and 4, and is 97% for the data shown in Fig. 3. After a period of variable duration, we perform quantum state tomography by applying either rotations about the x and y axes, or no rotation followed by a second 800 ns readout. Tomography results were corrected for the readout fidelity of 95%.

Calibration of the measurement. We calibrate the characteristic measurement time τ by examining histograms of the measurement signal for the qubit prepared in either the $|0\rangle$ or the $|1\rangle$ state. We prepare these states through a herald readout and then digitize the measurement signal for a variable period of time. The resulting distributions are approximately Gaussian:

$$P(V|0) = \sqrt{\frac{1}{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(V-\Delta V/2)^2}$$

$$P(V|1) = \sqrt{\frac{1}{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(V+\Delta V/2)^2}$$

We fit the distributions to determine ΔV , the voltage separation of the peaks and the variance σ^2 . The quantity $S = \Delta V^2/\sigma^2$ increases linearly with integration time: $S = 4t/\tau$. We fit this relationship to determine the characteristic measurement time τ .

To calibrate the initial state and the total dephasing rate, we prepare the qubit along the x axis and perform quantum state tomography after a variable period of time. The tomography results for the full ensemble are shown in Fig. 2a, and exhibit exponential decay of coherence at rate Γ . The total quantum measurement efficiency is given by $\eta_{\text{tot}} = 1/2\Gamma\tau$. Note that the total quantum measurement efficiency $\eta_{\text{tot}} = \eta_{\text{col}}\eta_{\text{amp}}\eta_{\text{env}}$ is the product of the efficiencies for collection, for amplification and from extra environmental dephasing. We use the tomography value at $t = 0$ to determine the initial state, denoted (x_0, z_0) .

To determine the Rabi frequency, Ω , we examine the ensemble tomography results as shown in Fig. 2b, c. The ensemble evolution is given by the Lindblad equation with arbitrary Rabi drive: $\dot{x}(t) = -\Gamma x(t) + \Omega z(t)$, $\dot{z}(t) = -\Omega x(t)$. With initial state (x_0, z_0) , these equations have an analytic solution

$$\begin{aligned} x(t) &= e^{-\Gamma t/2} \left(x_0 \cos(\lambda t) - \frac{\Gamma x_0 - 2\Omega z_0}{2\lambda} \sin(\lambda t) \right) \\ z(t) &= e^{-\Gamma t/2} \left(z_0 \cos(\lambda t) + \frac{\Gamma z_0 - 2\Omega x_0}{2\lambda} \sin(\lambda t) \right) \end{aligned} \quad (5)$$

where $\lambda = \sqrt{\Omega^2 - (\Gamma/2)^2}$. We use equation (5) to determine the Rabi frequency Ω for each measurement strength and Rabi drive amplitude.

Propagation of the qubit-state density matrix. Given the Rabi frequency Ω , the coherence decay rate γ and the initial qubit state calculated from the values of x_0 and z_0 at time $t = 0$, we propagate the initial state to states at later time steps $t = dt, 2dt, \dots, ndt$ using a two-step procedure. At any time t , we first apply a unitary rotation to account for the Rabi drive

$$\rho'_{01} = \rho_{01} + \frac{\Omega}{2} (\rho_{00} - \rho_{11}) dt \quad (6)$$

$$\rho'_{11} = \rho_{11} + \frac{\Omega}{2} (\rho'_{01} + \rho'_{10}) dt \quad (7)$$

where $\rho_{00}, \rho_{01}, \rho_{10}$ and ρ_{11} are matrix elements of a qubit density matrix $\rho(t)$. With the input values ρ'_{01} and ρ'_{11} , we next apply the Bayesian update to them based on the measurement result obtained in the time interval between t and $t + dt$, and get

$$\rho_{11}(t+dt) = \frac{(\rho'_{11}/\rho'_{00}) \exp(-4V(t)dt/\tau\Delta V)}{1 + (\rho'_{11}/\rho'_{00}) \exp(-4V(t)dt/\tau\Delta V)} \quad (8)$$

$$\rho_{01}(t+dt) = \rho'_{01} \frac{\sqrt{(1-\rho_{11}(t+dt))\rho_{11}(t+dt)}}{\sqrt{(1-\rho'_{11})\rho'_{11}}} e^{-\gamma dt} \quad (9)$$

We use $dt = 16$ ns as the data sampling interval, and $V(t)$ is the measurement result obtained between t and $t + dt$. As discussed in the main text, we validate the state update procedure using conditioned quantum state tomography and find good agreement between individual trajectories and the tomographic reconstructions.

Moreover, in the time-continuum limit $dt \rightarrow 0$, we can approximate the state update procedure (equations (6)–(9)) with the differential equations

$$\dot{x}(t) = -\gamma x(t) + \Omega z(t) - x(t)z(t)r(t)/\tau \quad (10)$$

$$\dot{z}(t) = -\Omega x(t) + (1 - z(t)^2)r(t)/\tau \quad (11)$$

where $r(t) = 2V(t)/\Delta V$ is the dimensionless measurement signal, and $x(t) = \text{tr}[\sigma_x \rho(t)]$ and $z(t) = \text{tr}[\sigma_z \rho(t)]$ are the Bloch vector coordinates as functions of time.

Tomographic validation. To verify that we have accurately tracked the quantum state of the system, we perform quantum state tomography at discrete times along the trajectory. We denote the target trajectory, which is based on a single run of the experiment, $(\tilde{x}(t), \tilde{z}(t))$. For each experimental sequence of total measurement duration t , we propagate ρ and, if $x(t) = \tilde{x}(t) \pm 0.03$ and $z(t) = \tilde{z}(t) \pm 0.03$, then the subsequent tomography results are included in the tomographic reconstruction of the state at time t . We repeat this analysis for all time steps between 0 and 1.6 μ s, and find good agreement between the individual trajectories and the tomographic reconstructions.

Histogram scaling. The greyscale histograms shown in Figs 3 and 4 represent the values of x and z at each time point, binning with the bin size of 0.02. The greyscale shading is normalized such that the most frequent value is 1 at each time point.

Derivation of the ordinary differential equations in equations (1)–(4). We consider a set of unitless measurement readouts $\{r_k\} = \{r_0, r_1, \dots, r_{n-1}\}$, where $r_k = 2V_k/\Delta V$, at times $\{t_k = kdt\}$ for $k = 0, 1, \dots, n-1$, and its corresponding set of qubit states, denoted by $\{q_k\}$. In our experiment, the y component of the qubit Bloch coordinates is always zero, and q_k is thus a two-dimensional vector: $q_k = (x_k, z_k)$. We write a joint probability density function of all measurement outcomes $\{r_k\}$, the quantum states $\{q_k\}$ and the chosen final state q_F , conditioned on the initial state q_I as

$$\begin{aligned} P(\{q_k\}, \{r_k\}, q_F | q_I) &= \delta^2(q_0 - q_I) \delta^2(q_n - q_F) \\ &\times \left(\prod_{k=0}^{n-1} P(q_{k+1} | q_k, r_k) P(r_k | q_k) \right) \end{aligned} \quad (12)$$

Here $P(q_{k+1} | q_k, r_k)$ is a probability density function of a qubit state at time t_{k+1} given a qubit state and measurement signal at previous time t_k . Because a qubit state at any time t_{k+1} is updated deterministically from q_k and r_k , the density function $P(q_{k+1} | q_k, r_k)$ is a delta function whose argument imposes the state update equations. The conditional distribution of the detector output $P(r_k | q_k)$ obtained in a time interval dt is a probability density function of r_k given q_k :

$$P(r_k | q_k) = \sqrt{\frac{dt}{2\pi\tau}} \left(\frac{1+z_k}{2} e^{-\frac{dt}{2\tau}(r_k-1)^2} + \frac{1-z_k}{2} e^{-\frac{dt}{2\tau}(r_k+1)^2} \right)$$

By expressing the delta functions in equation (12) in Fourier-transformed forms with conjugate variables $p_k = (p_k^x, p_k^z)$ for $k = -1, 0, \dots, n$ and other terms in exponential form, we can write the joint probability density function in a path integral representation $P(\{q_k\}, \{r_k\}, q_F | q_I) \propto \int \mathcal{D}p e^S$. Here $\mathcal{D}p$ is an integral measure over conjugate variables $\{p_k\}$, and S is an action given by

$$\begin{aligned} S &= -p_{-1} \cdot (q_0 - q_I) - p_n \cdot (q_n - q_F) \\ &+ \sum_{k=0}^{n-1} \{ -p_k \cdot (q_{k+1} - \mathcal{E}[q_k, r_k]) + \ln P(r_k | q_k) \} \end{aligned} \quad (13)$$

$$\begin{aligned} &= -B + \int_0^T dt [-p_x \dot{x} - p_z \dot{z} + p_x (-\gamma x + \Omega z - xzr/\tau) \\ &+ p_z (-\Omega x + (1 - z^2)r/\tau) \\ &- (r^2 - 2rz + 1)/2\tau] \end{aligned} \quad (14)$$

where we have used the operator $\mathcal{E}[q_k, r_k]$ to indicate the state update, and B as a short-hand for the first two terms in equation (13). We note that, in equation (14), we have taken the time-continuum limit $dt \rightarrow 0$ and written the action explicitly

for our qubit measurement case with the state update equations (10) and (11). We have also used shortened notation for the variables, for example $x = x(t) \equiv \lim_{dt \rightarrow 0} \{x_0, x_1, \dots, x_n\}$. To obtain the most likely path, we then extremize the action in equation (14) over all variables (x, z, p_x, p_z, r) and obtain the ordinary differential equations (ODEs) shown in equations (1)–(4):

$$\dot{x} = -\gamma x + \Omega z - xz r / \tau \quad (15)$$

$$\dot{z} = -\Omega x + (1 - z^2)r / \tau \quad (16)$$

$$\dot{p}_x = +\gamma p_x + \Omega p_z + p_x z r / \tau \quad (17)$$

$$\dot{p}_z = -\Omega p_x + (p_x x + 2p_z z - 1)r / \tau \quad (18)$$

Here $r = z + p_z(1 - z^2) - p_x x z$ and the forced boundary conditions are $x(t=0) = x_i, z(t=0) = z_i, x(t=T) = x_f, z(t=T) = z_f$. As discussed in the main text, we can analytically solve the ODEs in equations (15)–(18) when $\Omega = 0$. For the driven case, where $\Omega \neq 0$, we solve the equations numerically using a shooting method.

Interpretation of the solutions of the ODEs. Here we discuss the interpretation of the solution of the ODEs in equations (15)–(18) (and equations (1)–(4)). The extremization of the action in equation (13) can also be interpreted as a constrained optimization of the last term of equation (13), $\sum_{k=0}^{n-1} \ln P(r_k | q_k)$, which is the log-likelihood of the trajectory. The constraints are as follows: (1) the qubit state updates $q_{k+1} = \mathcal{E}[q_k, r_k]$ for $k = 0, 1, \dots, n-1$; (2) the pre-selected state is $q_0 = q_i$; and (3) the post-selected state is $q_n = q_f$. The conjugate variables $\{p_k\}$ now act as the Lagrange multipliers of the constrained optimization. With this interpretation, a solution of the ODEs in equations (15)–(18) therefore represents a path with an optimized value of $\sum_{k=0}^{n-1} \ln P(r_k | q_k)$ or its exponential $\prod_{k=0}^{n-1} P(r_k | q_k)$, that is, a measurement path probability density.

The most likely path. The optimized path mentioned in the previous subsection can represent either a maximum, a minimum or a saddle point of the path probability under the constraints. We can determine which by finding paths slightly varied from the optimized solution, with all constraints still applied. This can be done by adding small constants δ_1 and δ_2 to the right-hand sides of the differential equations of the conjugate variables p_x and p_z (equations (17) and (18)), leaving the equations for x and z unchanged, and solving the whole system with the same boundary conditions. Solutions of the modified ODEs will be slightly varied from the optimized path. We then compute their full-path probabilities, comparing with the probability calculated from the optimized path. In Extended Data Fig. 2, we show samples of paths from the variational method described here and the unnormalized full-path probability of the surrounding paths. In this case, it shows that the optimized solution is the most likely path, with a maximum value of the path probability density.

The most likely paths from the experimental post-selected trajectories. To find the most likely path from experimental trajectory data, we first define the closeness of any two trajectories (named a and b) as a time-average of the Euclidean distance:

$$D \equiv (1/n) \sum_{k=0}^{n-1} \sqrt{(x_a(t_k) - x_b(t_k))^2 + (z_a(t_k) - z_b(t_k))^2}$$

where $t_k = kdt$ and dt is the time step. For a qubit, the Euclidean distance is the trace distance (up to a factor of 2) between two quantum states ρ_1 and ρ_2 , given by $(1/2)\text{tr}[\rho_1 - \rho_2]$. We compute the distance D between all possible pairs of trajectories starting from the same initial state $q_i = (x_i, z_i)$ and ending around a final state $q_f = (x_f, z_f)$ with a small tolerance. We then search for N trajectories that have minimum average distance to all others, and average them to obtain an estimate of the most likely path. The number of trajectories N is chosen to be about 10% of the total number of trajectories in the sub-ensemble. (N is of order 10^2 , compared with 10^3 trajectories in the sub-ensemble). As a result, we get a smooth estimate of the most likely path, which is still very different from the total (sub-ensemble) averaged trajectory. We plot the standard deviation of the data of x and z for the chosen 10% of trajectories for every time step as a shaded band in Figs 3 and 4. As shown in those figures, the experimentally determined most likely paths closely approximate the theoretical most likely paths, that is, the solutions of the ODEs in equations

(15)–(18). We expect that the approximated curves converge to the smooth theory curves in the limit of an infinite ensemble of post-selected trajectories.

In some cases, we can simply look at a trajectory of local medians (medians of x or z at all time steps) and compare it with the theoretical most likely path. The median trajectory can in practice be a good approximation to the theory curve when the distribution of the post-selected trajectories is a narrow band, that is, when the post-selected trajectories lie closely around a single path. As an example, in the case where there is no drive on the qubit, $\Omega = 0$, we show in our theory paper⁶ that the median curves agree quite well with the most likely curves. However, in the driven case where the qubit trajectories can possibly have different winding numbers around the y axis, resulting in multiple most likely paths from an initial state to a final state, simply finding the median of the distribution of x or z is not enough to capture their most likely behaviour. In this paper, we focus only on the cases where there is a single most likely path between any two boundary states. We will discuss our findings concerning the multiple paths connecting two boundary states in a future work.

The most likely time. Apart from the path of maximum likelihood taken between the pre- and post-selected states in a fixed time, a complementary problem in quantum control is that of the optimal waiting time between starting and destination states. In the case where there is no Rabi drive on the qubit, $\Omega = 0$, we can fix the states at the endpoints and inquire about the most likely time taken to travel between them. While a path integral derivation of the most likely time is possible, we give a simpler derivation here based on the probability distribution of the time-average measurement readout $V = (1/n) \sum_{k=0}^{n-1} V_k$.

In the case with no drive on the qubit, the z coordinate of the qubit on the Bloch sphere at any time T is solely determined by V . We can derive the distribution of the final z coordinate (z_f) at any time T , given the initial z coordinate (z_i), $P(z_f | z_i)$, from the probability density function $P(V | z_i)$. The probability density function of V , given z_i , is

$$P(V | z_i) = P(V | 0) \frac{1 + z_i}{2} + P(V | 1) \frac{1 - z_i}{2} \\ = \sqrt{\frac{1}{2\pi\sigma^2}} \left(\frac{1 + z_i}{2} e^{-\frac{1}{2\sigma^2}(V - \Delta V/2)^2} + \frac{1 - z_i}{2} e^{-\frac{1}{2\sigma^2}(V + \Delta V/2)^2} \right)$$

where the variance of the voltage signal measured in a time dt is $\sigma^2 = \Delta V^2 \tau / 4dt$. We change variable from the time-averaged measurement signal V to the final z component z_f as follows: $V = (\tau \Delta V / 2T) [\tanh^{-1}(z_f) - \tanh^{-1}(z_i)]$. We obtain the differential measure $dV = (\tau / 2T) [\Delta V / (1 - z_f^2)] dz_f$. The probability density function of z_f given z_i can be computed via the relation $P(z_f | z_i) dV = P(V | z_i) dz_f$:

$$P(z_f | z_i) = \frac{\sqrt{\tau/2\pi T}}{(1 - z_i^2)} \times \exp \left\{ -\frac{T}{2\tau} (\bar{r}^2 + 1) + \frac{1}{2} \ln \left(\frac{1 - z_i^2}{1 - z_f^2} \right) \right\}$$

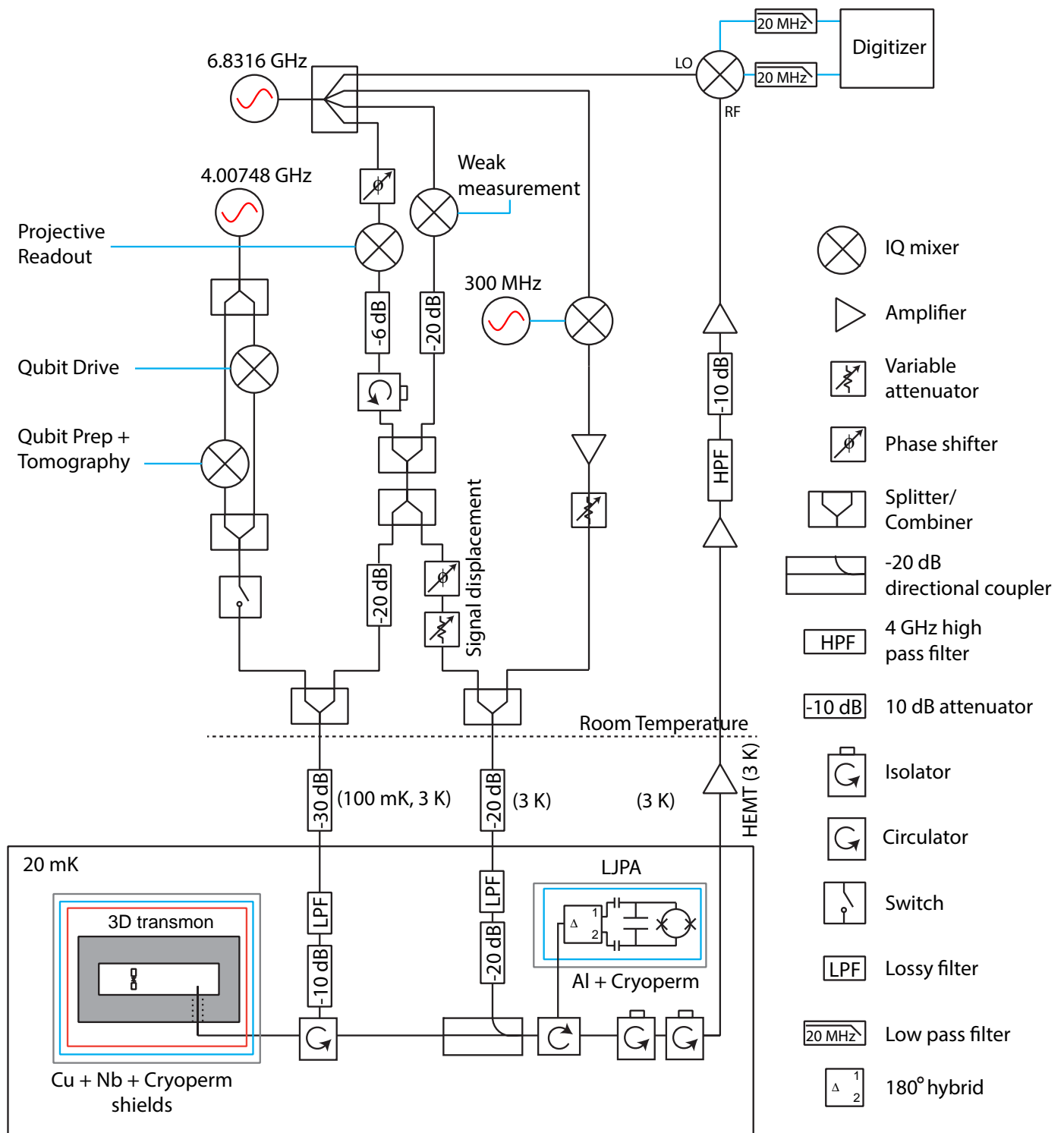
where $\bar{r} \equiv \frac{\tau}{T} \tanh^{-1} \left(\frac{z_f - z_i}{1 - z_i z_f} \right) = \frac{\tau}{T} (\tanh^{-1}(z_f) - \tanh^{-1}(z_i))$. For the case where the initial state is $x = +1$ ($z_i = 0$), the probability density function simplifies to

$$P(z_f | z_i = 0) = \frac{\sqrt{\frac{\tau}{2\pi T}}}{(1 - z_f^2)^{3/2}} \exp \left\{ -\frac{T}{2\tau} - \frac{\tau}{2T} [\tanh^{-1}(z_f)]^2 \right\}$$

We then compute the most likely time T_{opt} where the probability density function $P(z_f | z_i)$ is maximized for the fixed values of z_i and z_f . By maximizing the probability function $P(z_f | z_i)$ with respect to T , we obtain

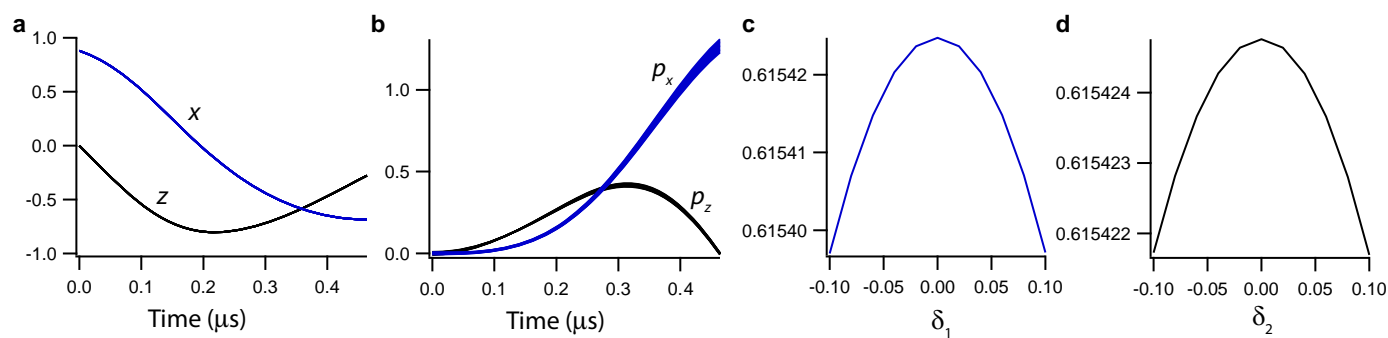
$$T_{\text{opt}} = \tau \left(\frac{\sqrt{1 + 4\bar{r}^2} - 1}{2} \right)$$

where $\bar{r} \equiv \tanh^{-1} \left(\frac{z_f - z_i}{1 - z_i z_f} \right)$. We show in Extended Data Fig. 3 the distributions $P(z_f | z_i = 0)$ as a function of time T for $z_f = 0.2, 0.4, 0.6$. They show very good agreement with the experimental data.



Extended Data Figure 1 | Experimental schematic. The weak measurement tone is always on. The projective readout tone is pulsed. The amplitude and phase of the signal displacement tone are adjusted to displace the measurement

signals back to the origin of the X_1 – X_2 plane, which allows the LIPA to perform in the linear regime.

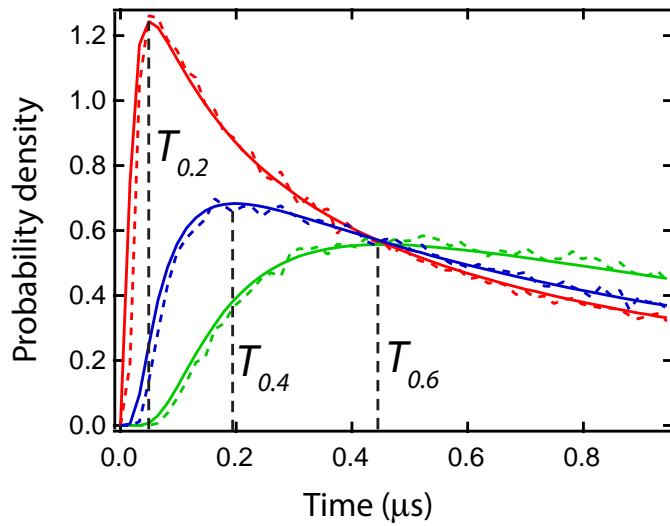


Extended Data Figure 2 | Paths slightly varied from the optimal solution.

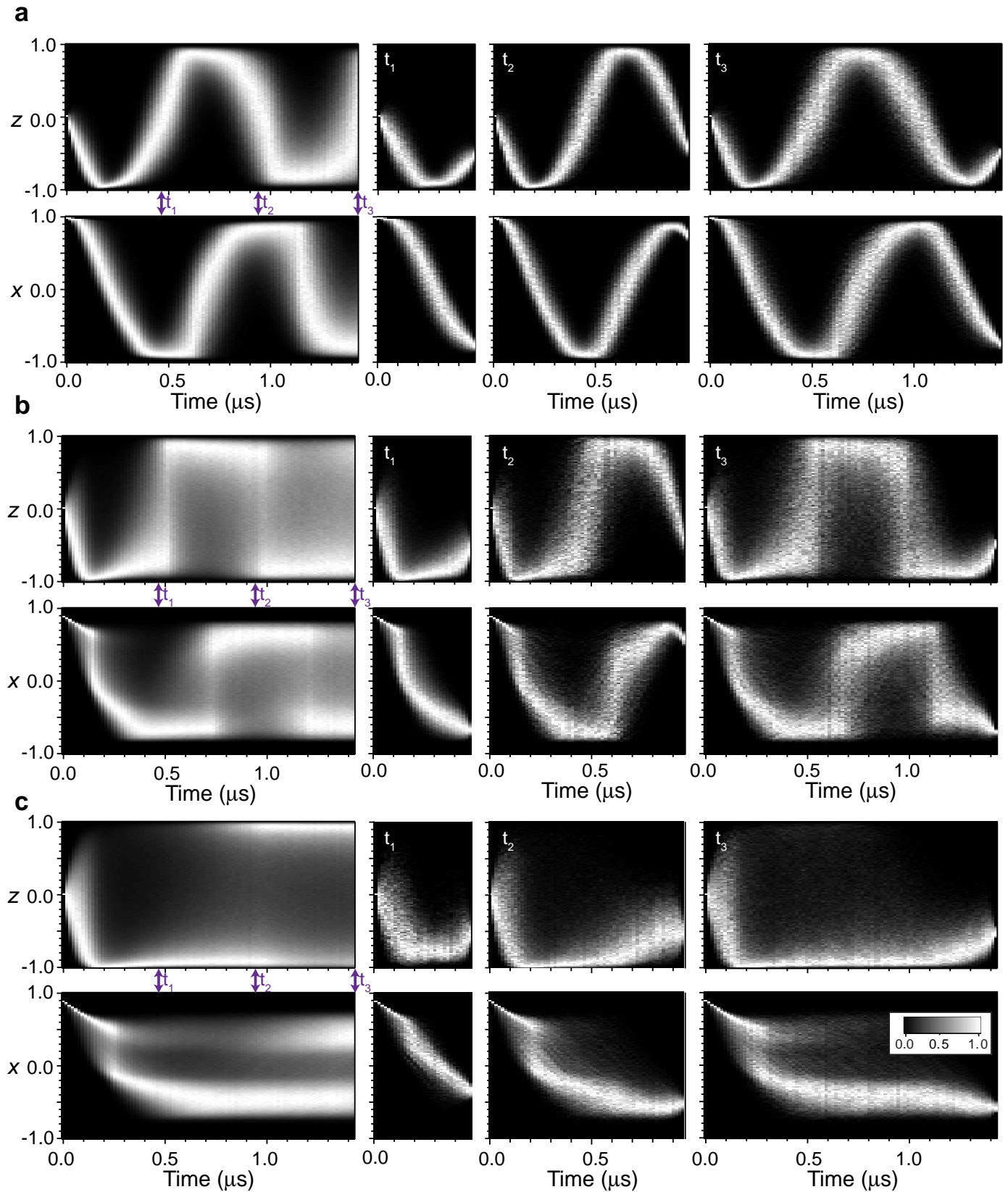
a, Overplotted x and z coordinates of 11 trajectories slightly varied from an optimized solution with boundary conditions $(x_i, z_i) = (0.88, 0)$, $(x_F, z_F, T_F) = (-0.683, -0.227, 0.464 \mu\text{s})$ and the Rabi drive $\Omega/2\pi = 1.08 \text{ MHz}$.

b, The corresponding conjugate variables p_x and p_z . **c**, **d**, Plots of the

unnormalized probability versus changes of the constant δ_1 in the p_x differential equation (**c**) and the unnormalized probability versus changes of the constant δ_2 in the p_z differential equation (**d**). In this case, the optimized solution gives a maximum value of the path probability density.



Extended Data Figure 3 | Optimal time between starting and destination states. The probability density functions $P(z_F|z_I = 0)$ plotted as functions of time T (solid curves) along with experimental data (dotted curves) with $\tau = 1.25 \mu\text{s}$. The red, green and blue curves are the distribution functions $P(z_F = 0.2|z_I = 0)$, $P(z_F = 0.4|z_I = 0)$ and $P(z_F = 0.6|z_I = 0)$, respectively. The optimized times T_{opt} for the three cases are shown as the vertical black dashed lines with the labels $T_{0.2}$, $T_{0.4}$ and $T_{0.6}$.



Extended Data Figure 4 | Greyscale histograms of ensemble and post-selected trajectories for different Rabi frequencies and measurement strengths. **a**, Ensemble and post-selected trajectories for $\Omega/2\pi = 1.08$ MHz and $\tau = 1.25$ μs . The post-selections for times $\{t_1 = 464$ ns, $t_2 = 944$ ns, $t_3 = 1.424$ $\mu\text{s}\}$ are $(x_F, z_F) = \{(-0.78, -0.5), (0.7, -0.5), (-0.73, -0.5)\}$ with a post selection window of ± 0.08 . **b**, Trajectories for $\Omega/2\pi = 1.08$ MHz and

$\tau = 315$ ns with $(x_F, z_F) = \{(-0.69, -0.5), (0.5, -0.5), (-0.73, -0.5)\}$. **c**, Trajectories for $\Omega/2\pi = 0.58$ MHz and $\tau = 315$ ns with $(x_F, z_F) = \{(-0.35, -0.5), (-0.5, -0.5), (-0.56, -0.5)\}$. Note that all the trajectories use the same value of z_F . The values of x_F were chosen to give a large number of trajectories in the post-selected ensemble.

Antarctic glaciation caused ocean circulation changes at the Eocene–Oligocene transition

A. Goldner^{1,2}, N. Herold³ & M. Huber^{3,4}

Two main hypotheses compete to explain global cooling and the abrupt growth of the Antarctic ice sheet across the Eocene–Oligocene transition about 34 million years ago: thermal isolation of Antarctica due to southern ocean gateway opening^{1–4}, and declining atmospheric CO₂ (refs 5, 6). Increases in ocean thermal stratification and circulation in proxies across the Eocene–Oligocene transition have been interpreted as a unique signature of gateway opening^{2,4}, but at present both mechanisms remain possible. Here, using a coupled ocean–atmosphere model, we show that the rise of Antarctic glaciation, rather than altered palaeogeography, is best able to explain the observed oceanographic changes. We find that growth of the Antarctic ice sheet caused enhanced northward transport of Antarctic intermediate water and invigorated the formation of Antarctic bottom water, fundamentally reorganizing ocean circulation. Conversely, gateway openings had much less impact on ocean thermal stratification and circulation. Our results support available evidence that CO₂ drawdown—not gateway opening—caused Antarctic ice sheet growth, and further show that these feedbacks in turn altered ocean circulation. The precise timing and rate of glaciation, and thus its impacts on ocean circulation, reflect the balance between potentially positive feedbacks (increases in sea ice extent and enhanced primary productivity) and negative feedbacks (stronger southward heat transport and localized high-latitude warming). The Antarctic ice sheet had a complex, dynamic role in ocean circulation and heat fluxes during its initiation, and these processes are likely to operate in the future.

Two main conceptual models have been advanced to explain the pattern of cooling⁷, enhanced Antarctic circumpolar circulation¹, biospheric transitions⁸ and glaciation across the Eocene–Oligocene transition (EOT) (~34.1–33.6 Myr ago). These are Southern Ocean gateway opening and CO₂ drawdown. The first model, supported by proxy interpretations and numerical ocean modelling, posits that opening of Southern Ocean gateways caused a fundamental reorganization of ocean structure and circulation^{1–4}, cooled the Antarctic region¹ and enhanced ocean heat transport into the North Atlantic^{9,10}. The second model is that CO₂ drawdown and its radiative forcing⁵ led to global cooling, enhanced near the poles, which initiated Antarctic glaciation. This conceptual model is supported by clear evidence of drawdown across the EOT^{6,11}, as well as by modelling of ice sheets^{5,12} and of the coupled ocean–atmosphere^{13,14}.

Although both conceptual models are supported by data and numerical models, both hypotheses have deficiencies. The gateway model explains many features of the ocean structure and circulation but has three main weaknesses. First, the timing of gateway opening is poorly constrained¹⁵ and does not seem to match the formation of the Antarctic ice sheet (AIS). Second, gateway opening is too gradual to explain rapid changes easily¹⁶, although nonlinear, thresholded responses cannot be ruled out. Third, most modelling suggests that open southern high-latitude gateways would have little impact on AIS formation^{5,12}.

The CO₂ drawdown argument also has weaknesses. First, it does not explain the ocean thermal stratification in benthic $\delta^{18}\text{O}$ records as the gateway mechanism does^{2,4}. Second, the magnitude of the decrease

in CO₂ required to match reconstructed cooling in some previous fully coupled EOT modelling⁷ is at least twice as large as CO₂ proxies indicate.

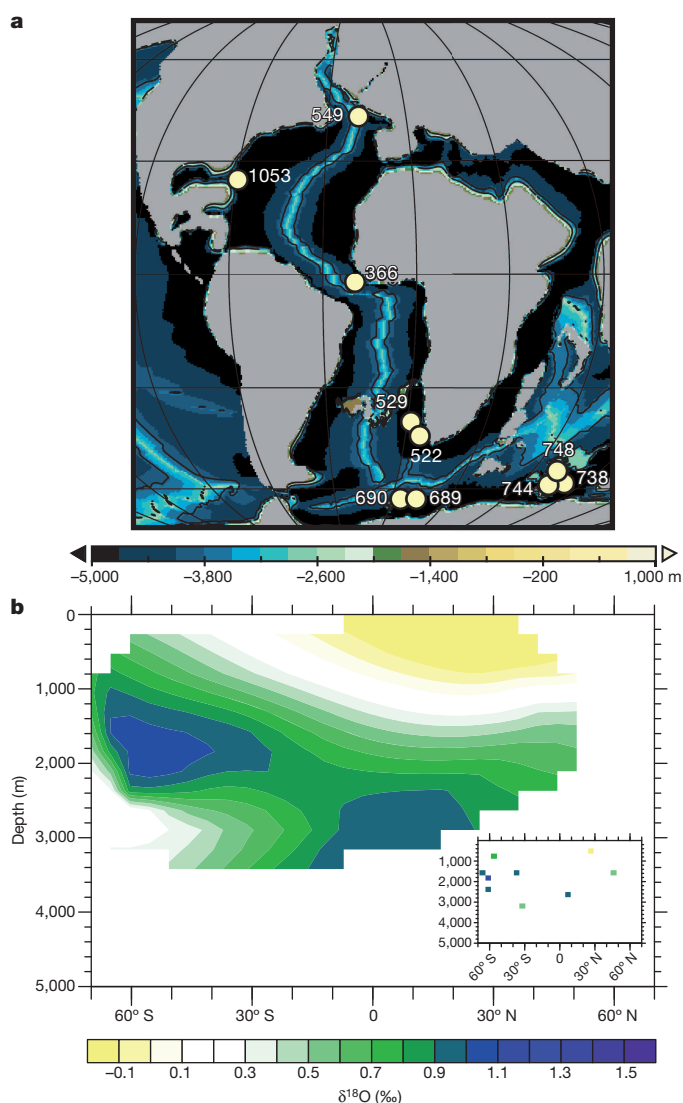


Figure 1 | Atlantic Ocean $\delta^{18}\text{O}$ records compiled for the EOT. **a**, Deep Sea Drilling Project (DSDP) and Ocean Drilling Program (ODP) site locations for $\delta^{18}\text{O}$ records (palaeorecords described in Extended Data Table 1) overlaid on late Eocene bathymetry. **b**, Positive values indicate an increase in $\delta^{18}\text{O}$ going from the late Eocene (34–36 Myr ago) into the early Oligocene (31–29 Myr ago) interpolated across the Atlantic Ocean basin (model longitudes 315–359°). Raw $\delta^{18}\text{O}$ anomalies are plotted as the inset in **b** (same values as in Extended Data Fig. 9a).

¹Department of Earth, Atmospheric, and Planetary Sciences, Purdue University, West Lafayette, Indiana 47907, USA. ²American Geophysical Union, Washington DC 20009, USA. ³Department of Earth Sciences, University of New Hampshire, Durham, New Hampshire 03824, USA. ⁴Earth Systems Research Center, Institute for Earth, Ocean and Space Sciences, University of New Hampshire, Durham, New Hampshire 03824, USA.

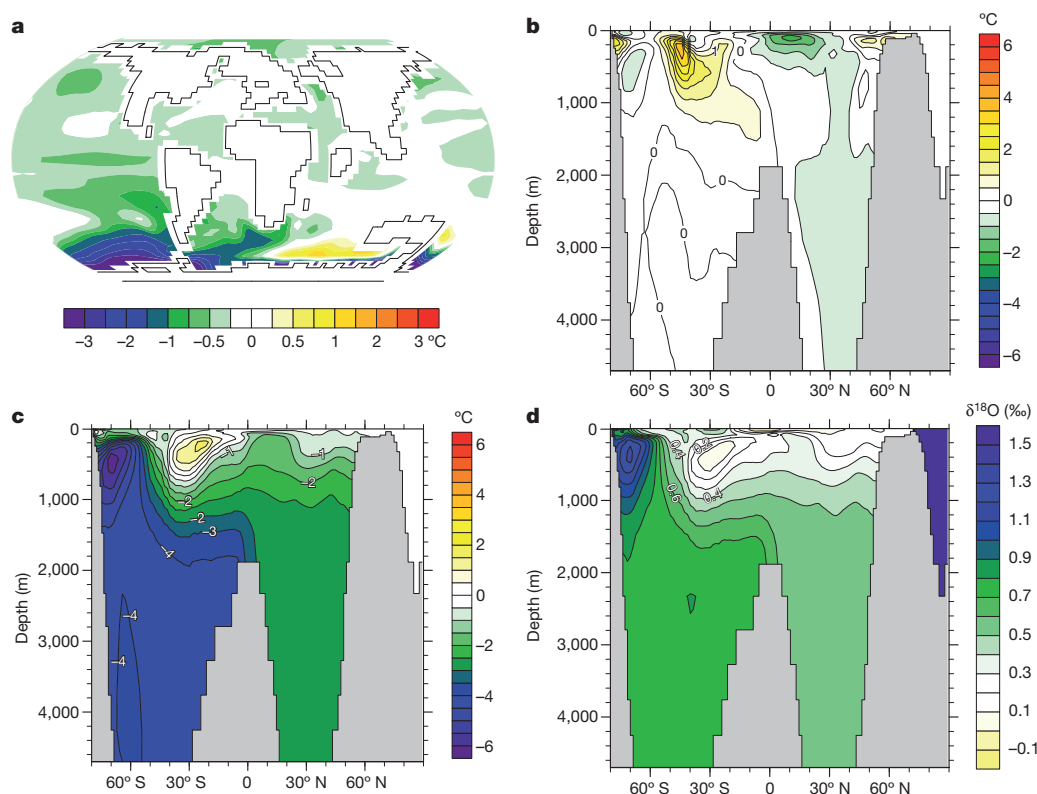


Figure 2 | Ocean temperature and $\delta^{18}\text{O}$ anomalies. **a**, Zonally averaged ocean temperature anomalies induced by glaciation with CO_2 held fixed at 1,120 p.p.m. **b**, Ocean temperature anomaly in the Atlantic Ocean basin induced by opening Southern Ocean gateways with CO_2 held fixed at 1,120 p.p.m. **c**, As **a**, but averaged over the Atlantic Ocean. **d**, Estimated Atlantic Ocean $\delta^{18}\text{O}$ anomaly induced by glaciation.

Recent work extended the CO_2 drawdown hypothesis by suggesting that AIS formation, caused by a decrease in greenhouse gases, was responsible for strengthening the ocean's overturning circulation through enhanced sea-ice growth and brine rejection¹⁷. This is conjectured to have invigorated the formation rates of Antarctic bottom water, explaining changes in the oceans thermal gradient detected across the EOT. No coupled ocean–atmosphere modelling studies have been performed to test the hypothesis that growth of the AIS during the EOT could explain reconstructed changes in ocean temperatures, $\delta^{18}\text{O}$ gradients and circulation.

In this study we show that inclusion of this simple feedback—AIS formation—is enough to bring the CO_2 drawdown hypothesis into better agreement with both enhanced gradients noted in ocean $\delta^{18}\text{O}$ records and observed cooling at the correct CO_2 (see Methods and Extended Data Fig. 1). First we review this isotopic signature and then use an Earth system model to show that these results can be reconciled.

The pattern of global surface cooling is now well established from many independent proxy records^{7,18–20}, but cooling inferred from proxies is not homogeneous. Annual mean temperature decreases weakly across the EOT in northern high latitudes and in the subtropics¹⁸, whereas cooling is strong in the high southern latitudes⁷. This information has been used in conjunction with benthic $\delta^{18}\text{O}$ records^{2,4} to estimate that the EOT is associated with a deep ocean cooling by several degrees^{18,19} and the growth of an AIS to 50–100% of the modern volume⁷. This inhomogeneous signature has been proposed to be uniquely explained by Southern Ocean gateway opening^{2,4,9}.

To illustrate this signature, we recompile existing Atlantic Ocean benthic foraminiferal $\delta^{18}\text{O}$ records (Fig. 1a and Extended Data Table 1 and Extended Data Fig. 9a) and use natural neighbour regridding to generate a south–north transect along the Atlantic Ocean basin (Fig. 1b). As has been pointed out previously^{2,4}, isotope changes across the EOT are consistent with abyssal cooling, but the data support weak warming or negligible cooling in the upper subtropical ocean; that is, the vertical and meridional gradients in $\delta^{18}\text{O}$ and presumably temperature were enhanced across the EOT. Opening of Southern Ocean gateways, which initiated the Drake Passage (DP) effect⁹, is one potential cause of this signature, whereas CO_2 drawdown by itself does not reproduce this pattern⁷. Here

we test the hypothesis that growth of the AIS, a feedback ignored in previous work, might produce this signature as well.

We employ a fully coupled atmosphere, sea-ice, land and ocean model with representative EOT boundary conditions to quantify the changes in ocean temperature, salinity and circulation that occur as a result of adding the AIS into the unglaciated late Eocene world. We perform a series of simulations using the National Center for Atmospheric Research Community Earth System Model (CESM1.0). The unglaciated EOT simulations employ topography, vegetation, modern orbital variations, bathymetry and ocean gateways described in previous work²¹. For the glaciated EOT simulations, we increase topography and albedo over Antarctica to nearly modern-day levels ($\sim 75\%$ of modern Antarctic ice volume).

The simulations were integrated for 3,500 model years, beginning from the end of previously described equilibrated simulations with an earlier model version. Equilibration was verified after calculating near-zero surface and deep-ocean trends in temperature and ideal-age tracers (see Methods). In the first experiment we add the AIS while holding CO_2 constant at 1,120 p.p.m. In the second experiment CO_2 is decreased from 1,120 to 560 p.p.m. while keeping Antarctica unglaciated. In the third experiment, we combine AIS and CO_2 perturbations to explore the combined impact on the ocean's structure. Finally, we explore the impact of the DP and the Tasman Gateway (TG) by opening and closing both high-latitude gateways, holding everything else fixed.

First, we compare the glaciated and unglaciated EOT simulations at constant CO_2 to understand the response to glaciation alone. We find that AIS growth engenders a complex response of regional warming and cooling consistent with a previous related study in which ocean heat transport was specified²². Subtle warming occurs at the surface in the Southern Indian Ocean and within subtropical gyres, but southern high latitudes cool, especially in the Atlantic, and this is communicated into the abyss (Fig. 2a, c and Extended Data Fig. 1). Glaciation shifts the ocean's meridional and vertical thermal structure as the deep ocean cools more in the Southern Hemisphere than in the Northern Hemisphere, and cooling above the thermocline is negligible whereas cooling below it is strong. This glaciation-induced cooling is similar to that predicted from opening Southern Ocean gateways in previous work that did not use fully coupled

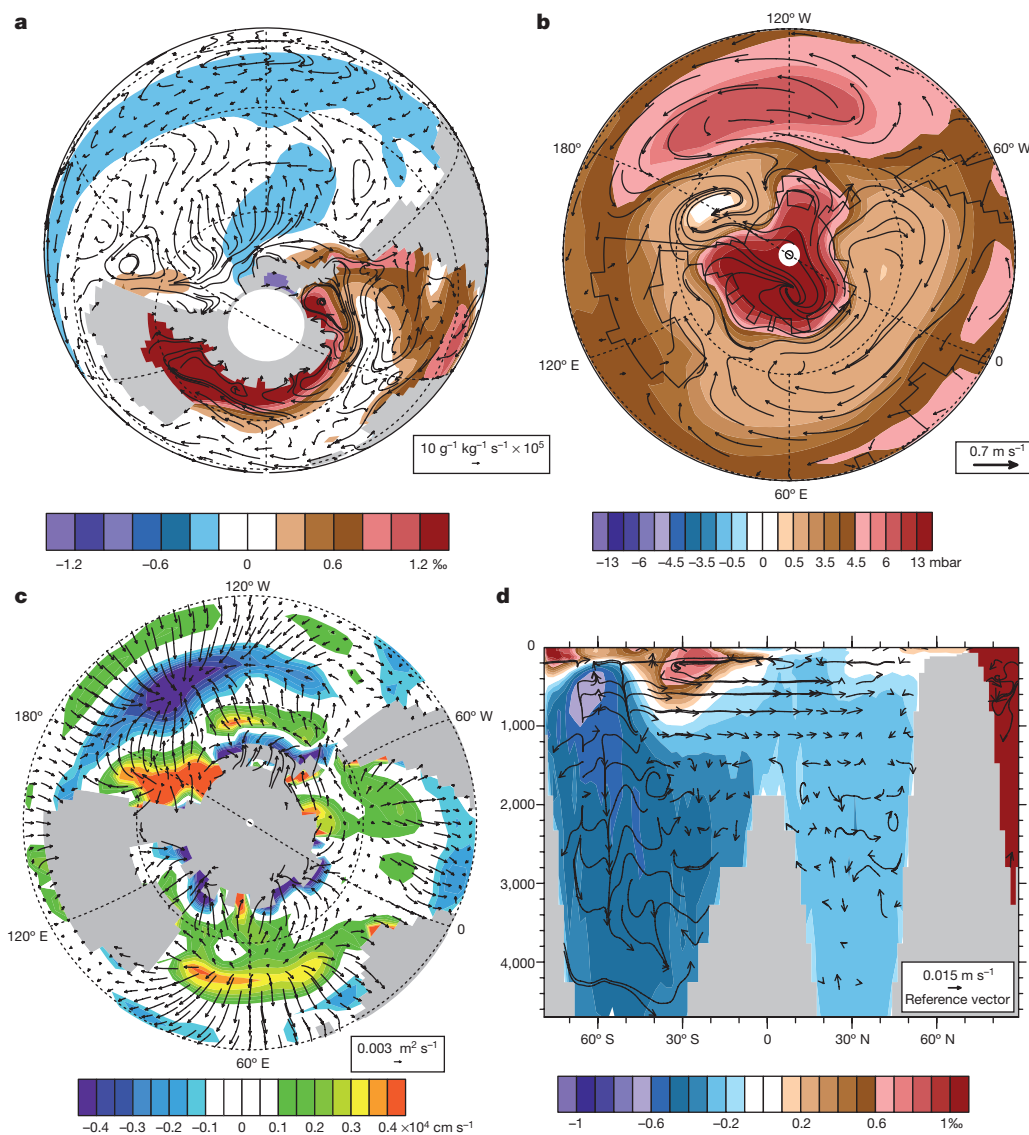


Figure 3 | Diagnostics for the mechanism for deep ocean cooling due to glaciation. **a**, Anomalous salinity (colour contour) and surface salt flux (vectors). **b**, Sea-level pressure anomaly (colour contour) and mean surface wind anomaly (vectors). **c**, Anomalous upwelling and downwelling (colour contour) and Ekman transport (vectors). **d**, Zonally averaged salinity (colour contour) and ocean current anomalies in Atlantic Ocean (vectors). (See Extended Data Figs 4–7 for absolute fields, including sea level pressure, surface salt flux, Ekman transport and ocean velocities.)

models²³. With a coupled model we simulate much less cooling due to Southern Ocean gateway opening (Fig. 2b). The pattern of cooling induced by glaciation has interesting implications for interpreting $\delta^{18}\text{O}$ records.

To facilitate comparison with $\delta^{18}\text{O}$ records we use an empirical calibration between temperature and salinity to express ocean model results as an equivalent $\delta^{18}\text{O}$ change (see Methods). Comparison between the model and data reveals a good match (Fig. 2d) in terms of the change in $\delta^{18}\text{O}$ gradient. Specifically, there is the positive isotopic shift in $\delta^{18}\text{O}$ in the southern high latitudes^{2,4}, and a relatively small change in $\delta^{18}\text{O}$ in the Northern Hemisphere. The $\delta^{18}\text{O}$ anomaly induced by opening ocean gateways does not produce a positive isotopic shift in the southern high latitudes (Extended Data Fig. 2). This is important because previous studies link the positive isotopic $\delta^{18}\text{O}$ excursion in Southern Hemisphere near the EOT to opening of the DP, and here we show that AIS growth is a different mechanism for shifting the ocean's $\delta^{18}\text{O}$ gradient.

The change in the meridional temperature gradient and deep ocean cooling due to glaciation is associated with circulation shifts in the atmosphere and ocean. Introduction of the AIS lowers sea surface temperatures (SSTs; Extended Data Fig. 3) and enhances the growth of sea ice along the Antarctic margin (Extended Data Fig. 3). These cooler Antarctic coastal waters are also saltier because an enhanced salt flux towards Antarctica within the Ekman layer (Fig. 3a and Extended Data Fig. 4) weakens the pycnocline. These changes occur because a shift in regional atmospheric circulation is induced by Antarctic glacial topography (Fig. 3b), which intensifies the pressure gradient between Antarctica and

the surrounding ocean. The change in pressure gradient drives increases in surface wind around Antarctica (Fig. 3b and Extended Data Fig. 5), enhancing Ekman transport towards Antarctica (Fig. 3c and Extended Data Fig. 6). Surface currents then subduct along the Antarctic coastline (Fig. 3d and Extended Data Fig. 7), moving the cooler, denser water mass into the abyss. Downwelling of high-salinity water increases at 60° S below the mixed layer, which contributes to enhanced deep-water formation. These changes in oceanic and atmospheric circulation in the southern high latitudes drive an enhanced meridional overturning circulation (Extended Data Fig. 8). Our simulated change in ocean circulation is consistent with increased northward transport of detected proxies of Antarctic intermediate water (Fig. 3d)⁴. Previous modelling studies also found increases in Ekman divergence and increased Atlantic deep water formation due to the opening of the DP⁹, although we find comparable shifts in circulation and ocean temperatures without changing gateways or CO_2 . These simulations reproduce many trends in ocean temperature and circulation produced in previous studies driven by CO_2 changes alone⁷, but inclusion of AIS formation as a feedback enables these shifts to occur with more realistic changes in CO_2 .

The changed ocean circulation and structure increase ocean heat transport by ~ 0.3 PW (Fig. 4) or $\sim 60\%$ of the value at 45–60° S. Previous work found that a 20% increase in ocean heat transport could significantly affect the timing of glacial inception and AIS stability⁵. Thus, glaciation induces a negative feedback by increasing southward ocean heat transport and heat fluxes along the Antarctic margin, potentially capable of significant

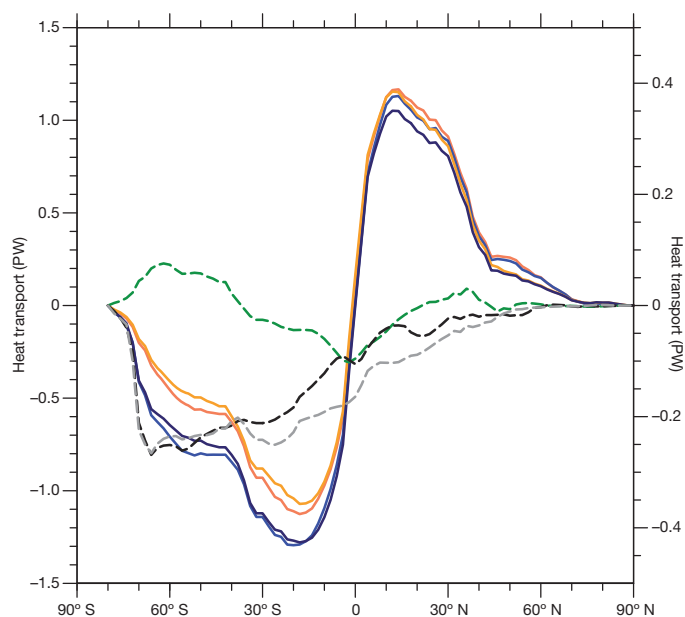


Figure 4 | Zonally averaged poleward ocean heat transport. Shown are unglaciated (orange line) and glaciated late Eocene (dark blue line) CO_2 specified at 1,120 p.p.m.; unglaciated (light orange line) and glaciated cases (light blue line) at 560 p.p.m. CO_2 ; anomaly (left y axis) due to glaciation at 1,120 p.p.m. CO_2 (dashed grey line) and 560 p.p.m. CO_2 (dashed black line); and anomaly (right y axis) due to gateway opening at 1,120 p.p.m. CO_2 (dashed green line).

feedback onto the AIS (Fig. 4). Such changes would also have significant impacts on other feedbacks, for example those due to sea ice^{12,24}. This negative feedback may help to explain the stepwise growth of the AIS⁵ and its early history of instability¹⁶. AIS formation is a leading-order feedback that should be considered in such studies. These results have implications for estimates of future changes in ice sheets and sea level, because the results demonstrate that ice sheets modify ocean stratification, casting doubt on the assumptions used in some studies²⁵. Our experiments opening the TG and DP gateways show weak changes in ocean heat transport (Fig. 4), in agreement with previous work in coupled models²⁶.

Consequently, although debate continues on whether the DP and TG opened shallowly in the Eocene or not^{15,27}, deep opening was probably not completed until the Miocene^{28,29}. Our simulations show little sensitivity to gateway opening and illustrate that it is unlikely that gateway opening is well suited to explain the rapid, large changes during the EOT (Fig. 2b and Extended Data Fig. 9). We have shown that the AIS–ocean feedback mechanism parsimoniously explains existing data and is consistent with reconstructed timing of events (Fig. 2). Our modelling results support the hypothesis that a decrease in CO_2 cooled the world and initiated AIS growth, which in turn increased the ocean's thermal gradient and invigorated circulation. Increases in wind-driven upwelling around Antarctica drove changes in ocean productivity seen in high-latitude records; they also enhanced ventilation, potentially drawing down carbon, and strengthened cooling across the EOT³⁰. Our results suggest that the net impact of glaciation was mediated by strong feedbacks, both negative and positive; a deeper understanding requires improved modelling and data integration. Nevertheless, one thing is clear: the AIS had a leading role in climate changes during glacial expansion across the EOT, and the same dynamics and feedbacks are likely to be important in the event of AIS retreat.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 15 April; accepted 19 June 2014.

1. Kennett, J. P. Cenozoic evolution of Antarctic glaciations, the circum-Antarctic ocean and their impact on global paleoceanography. *J. Geophys. Res.* **82**, 3843–3860 (1977).

2. Cramer, B. S., Toggweiler, J. R., Wright, J. D., Katz, M. E. & Miller, K. G. Ocean overturning since the Late Cretaceous: inferences from a new benthic foraminiferal isotope compilation. *Paleoceanography* **24**, PA4216 (2009).
3. Katz, M. E. *et al.* Stepwise transition from the Eocene greenhouse to the Oligocene icehouse. *Nature Geosci.* **1**, 329–334 (2008).
4. Katz, M. E. *et al.* Impact of Antarctic circumpolar current development on Late Paleogene ocean structure. *Science* **332**, 1076–1079 (2011).
5. DeConto, R. M. & Pollard, D. Rapid Cenozoic glaciation of Antarctica induced by declining atmospheric CO_2 . *Nature* **421**, 245–249 (2003).
6. Pagani, M. *et al.* The role of carbon dioxide during the onset of Antarctic glaciation. *Science* **334**, 1261–1264 (2011).
7. Liu, Z. *et al.* Global cooling during the Eocene–Oligocene climate transition. *Science* **323**, 1187–1190 (2009).
8. Houben, A. J. *et al.* Reorganization of Southern Ocean plankton ecosystem at the onset of Antarctic glaciation. *Science* **340**, 341–344 (2013).
9. Toggweiler, J. R. & Samuels, B. Effect of Drake Passage on the global thermohaline circulation. *Deep Sea Res. Oceanogr. Res. Pap.* **42**, 477–500 (1994).
10. Sijp, W. P., England, M. H. & Toggweiler, J. R. Effect of ocean gateway changes under greenhouse warmth. *J. Clim.* **22**, 6639–6652 (2009).
11. Pearson, P. N., Foster, G. L. & Wade, B. S. Atmospheric carbon dioxide through the Eocene–Oligocene climate transition. *Nature* **461**, 1110–1113 (2009).
12. Pollard, D. & DeConto, R. M. Hysteresis in Cenozoic Antarctic ice sheet variations. *Glob. Planet. Change* **45**, 9–21 (2005).
13. Huber, M. & Nof, D. The ocean circulation in the southern hemisphere and its climatic impacts in the Eocene. *Paleogeogr. Palaeoclimatol. Palaeoecol.* **231**, 9–28 (2006).
14. Lefebvre, V., Donnadieu, Y., Sepulchre, P., Swingedouw, D. & Zhang, Z. S. Deciphering the role of southern gateways and carbon dioxide on the onset of the Antarctic Circumpolar Current. *Paleoceanography* **27**, PA4201 (2012).
15. Lawver, L. A., Gahagan, L. M. & Dalziel, I. W. D. in *Tectonic, Climatic, and Cryospheric Evolution of the Antarctic Peninsula* (eds Anderson, J. B. & Wellner, J. S.) 5–33 (Am. Geophys. Un. Spec. Publ. 63, 2011).
16. Coxall, H. K., Wilson, P. A., Pälike, H., Lear, C. H. & Backman, J. Rapid stepwise onset of Antarctic glaciation and deeper calcite compensation in the Pacific Ocean. *Nature* **433**, 53–57 (2005).
17. Miller, K. G. *et al.* in *The Late Eocene Earth? Hothouse, Icehouse, and Impacts* (eds Koeberl, C. & Montanari, A.) 169–178 (Geol. Soc. Am. Spec. Pap. 452, 2009).
18. Lear, C. H., Bailey, T. R., Pearson, P. N., Coxall, H. K. & Rosenthal, Y. Cooling and ice growth across the Eocene–Oligocene transition. *Geology* **36**, 251–254 (2008).
19. Bohaty, S. M., Zachos, J. C. & Delaney, L. M. Foraminiferal Mg/Ca evidence for Southern Ocean cooling across the Eocene–Oligocene transition. *Earth Planet. Sci. Lett.* **317–318**, 251–261 (2012).
20. Hren, M. T. *et al.* Terrestrial cooling in Northern Europe during the Eocene–Oligocene transition. *Proc. Natl Acad. Sci. USA* **110**, 7562–7567 (2013).
21. Huber, M. & Caballero, R. The early Eocene equable climate problem revisited. *Clim. Past* **7**, 603–633 (2011).
22. Goldner, A., Huber, M. & Caballero, R. Does Antarctic glaciation cool the world? *Clim. Past* **9**, 173–189 (2013).
23. Sijp, W. P., England, M. H. & Huber, M. Effect of the deepening of the Tasman Gateway on the global ocean. *Paleoceanography* **26**, PA4207 (2011).
24. DeConto, R., Pollard, D. & Harwood, D. Sea ice feedback and Cenozoic evolution of Antarctic climate and ice sheets. *Paleoceanography* **22**, PA3214 (2007).
25. Foster, G. L. & Rohling, E. J. Relationship between sea level and climate forcing by CO_2 on geological timescales. *Proc. Natl Acad. Sci. USA* **110**, 1209–1214 (2013).
26. Hill, D. J. *et al.* Paleogeographic controls on the onset of the Antarctic circumpolar current. *Geophys. Res. Lett.* **40**, 5199–5204 (2013).
27. Bijl, P. K. *et al.* Eocene cooling linked to early flow across the Tasmanian Gateway. *Proc. Natl Acad. Sci. USA* **110**, 9645–9650 (2013).
28. Lyle, M., Gibbs, S., Moore, T. C. Jr & Rea, D. K. Late Oligocene initiation of the Antarctic circumpolar current: evidence from the South Pacific. *Geology* **35**, 691–694 (2007).
29. Dalziel, I. W. D. Drake Passage and the Scotia arc: a tortuous space–time gateway for the Antarctic Circumpolar Current. *Geology* **42**, 367–368 (2014).
30. Zachos, J. C. & Kump, L. R. Carbon cycle feedbacks and the initiation of Antarctic glaciation in the earliest Oligocene. *Global Planet. Change* **47**, 51–66 (2005).

Acknowledgements A.G. was funded by a Graduate Assistance in Areas of National Need (GAANN) fellowship through the Computational Sciences and Engineering Program at Purdue University. M.H. and N.H. were supported by National Science Foundation (NSF) P2C2 grants OCE 0902882 and EAR 1049921. Computing was performed on Rosen Center for Advanced Computing resources on the Hansen and Coates cluster at Purdue University. Proxy records were compiled from refs 2, 4, and specific records and references are described in extended data. The CESM model is supported and developed by the National Center for Atmospheric Research, which is supported by the NSF.

Author Contributions A.G. conducted the EOT simulations, recompiled the proxy record data and wrote the manuscript. N.H. and M.H. helped compile the proxy record data and helped with writing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.G. (aarongoldner@gmail.com) or M. H. (matthew.huber@unh.edu).

METHODS

Community Earth System Modelling (CESM) framework. CESM1.0 is a widely used, well-described and well-validated coupled model with atmosphere³¹, land³² and sea ice³³, and ocean model components. For the atmospheric component (CAM4) we use the spectral dynamical core at T31 resolution ($\sim 3.7^\circ \times 3.7^\circ$ per atmospheric grid cell) and the ocean model resolution is 122 latitudes by 100 longitudes ($\sim 1.8^\circ \times 3.6^\circ$ per ocean grid cell, with much finer resolution near the Equator and at high latitudes) with 25 unevenly spaced vertical levels in the ocean. The CESM ocean component (POP2) uses anisotropic horizontal viscosity³⁴, the Gent–McWilliams isopycnal mixing scheme³⁵ and an updated implementation of the KPP vertical mixing scheme³⁶, and has improved near-surface eddy flux parameterization, improving the effects of the diabatic mesoscale eddy fluxes^{37,38}, and improved restratification effects of the submesoscale mixed-layer eddies due to a revamped mixed-layer eddy parameterization³⁹. The simulation improvements compared with the previous version (CCSM3) include improved representation of the El Niño–Southern Oscillation (ENSO) and more realistic sea-ice coupling. The CESM framework is suitable for exploring EOT ocean circulation because it has proved capable of reproducing modern ocean currents and circulation⁴⁰ and we validated the main circulations produced in earlier versions of the model^{13,41}.

Experimental method. For the atmosphere we create aerosol forcing files specifically for the EOT; the description of this process is found in previous work⁴². We have prescribed global vegetation in the EOT simulations to be the same as in our earlier published studies^{21,41}, which restricts vegetative feedbacks. The ocean temperatures for the glaciated and unglaciated EOT simulations are initialized with a zonally averaged temperature distribution from the last 50 years of fully coupled CCSM3 Eocene modelling simulations run for more than 3,000 years with 560 and 1,120 p.p.m. CO₂ (refs 7, 41). The CESM1.0 fully coupled modelling simulations are then run for 3,400 years with the last 75 model years used for analysis. We conduct two glaciated Eocene simulations at 560 and 1,120 p.p.m. CO₂ and two unglaciated simulations at 560 and 1,120 p.p.m. CO₂. For the gateway simulations we branch two experiments from year 3,400 of the 1,120 p.p.m. CO₂ unglaciated Eocene simulation and run these experiments for a further >1,000 years. In the first gateway experiment we have both DP and TG open, and in the second gateway experiment we have the DP and TG closed.

The CO₂ range of 1,120–560 p.p.m. was chosen because this is close to the estimated change in atmospheric CO₂ across the EOT ($\sim 1,100$ p.p.m.) to lower levels around 600 p.p.m. after the transition^{6,11}. The volume of the Eocene AIS that we impose in the glaciation experiments is 20.3×10^6 km³, $\sim 75\%$ the size of the modern AIS. This is well within the estimated size for the AIS in the early Oligocene, for which estimated EOT ice volume is between 40% and 120% of modern^{7,43}. We diagnosed the surface mass balance of the imposed ice sheets and found them to be positive; that is, the simulations are self-consistent. When an ice sheet is imposed on Antarctica there is enough snow remaining in summer to keep the ice sheet intact.

Deep ocean equilibrium. To ensure that the coupled model simulations are in equilibrium, we employ a series of metrics. First, the surface energy balance is close to 0 in all simulations (about -0.04 W m⁻² in the unglaciated simulations and -0.06 W m⁻² in the glaciated simulations). Second, the volumetrically integrated ocean temperature trend from surface to 5 km is -0.016 °C per century over the last century in unglaciated simulations. This means that during a 100-year period the volumetrically integrated ocean temperature cooled by 0.016 °C. Third, a time series of surface and deep ocean temperature across the last 2,400 years of the simulation found no significant trend in the surface or deep ocean temperature. Fourth, visualizations of glaciated and unglaciated simulations plotting the temperature, salinity and ideal age trend for each simulation, demonstrate that the key patterns of interest in this study are established early and are stable through the length of the simulation.

Comparison with previous fully coupled EOT simulations. Previous coupled atmosphere–ocean EOT simulations required a CO₂ decrease of 2,240–560 p.p.m. (outside the range of proxy reconstructions) to match temperature proxy records^{7,20,44}. Here our change in CO₂ (1,120–560 p.p.m.) is more aligned with CO₂ levels supported by proxies and we find that the combined effect of glaciation and CO₂ (comparing the unglaciated 1,120 p.p.m. CO₂ case with the glaciated 560 p.p.m. CO₂ case) cools the surface and deep ocean ~ 4 °C, comparable to the surface cooling of 4.4 °C (Extended Data Fig. 1c), and benthic cooling of 4 °C described in previous fully

coupled modelling⁷. We also find a global mean surface temperature decrease of 4.3 °C, which compares well with the recent estimate of a 4–6 °C decrease in temperature across the EOT²⁰. These brief comparisons with previous work highlight the importance of including both AIS and CO₂ forcing in EOT climate simulations.

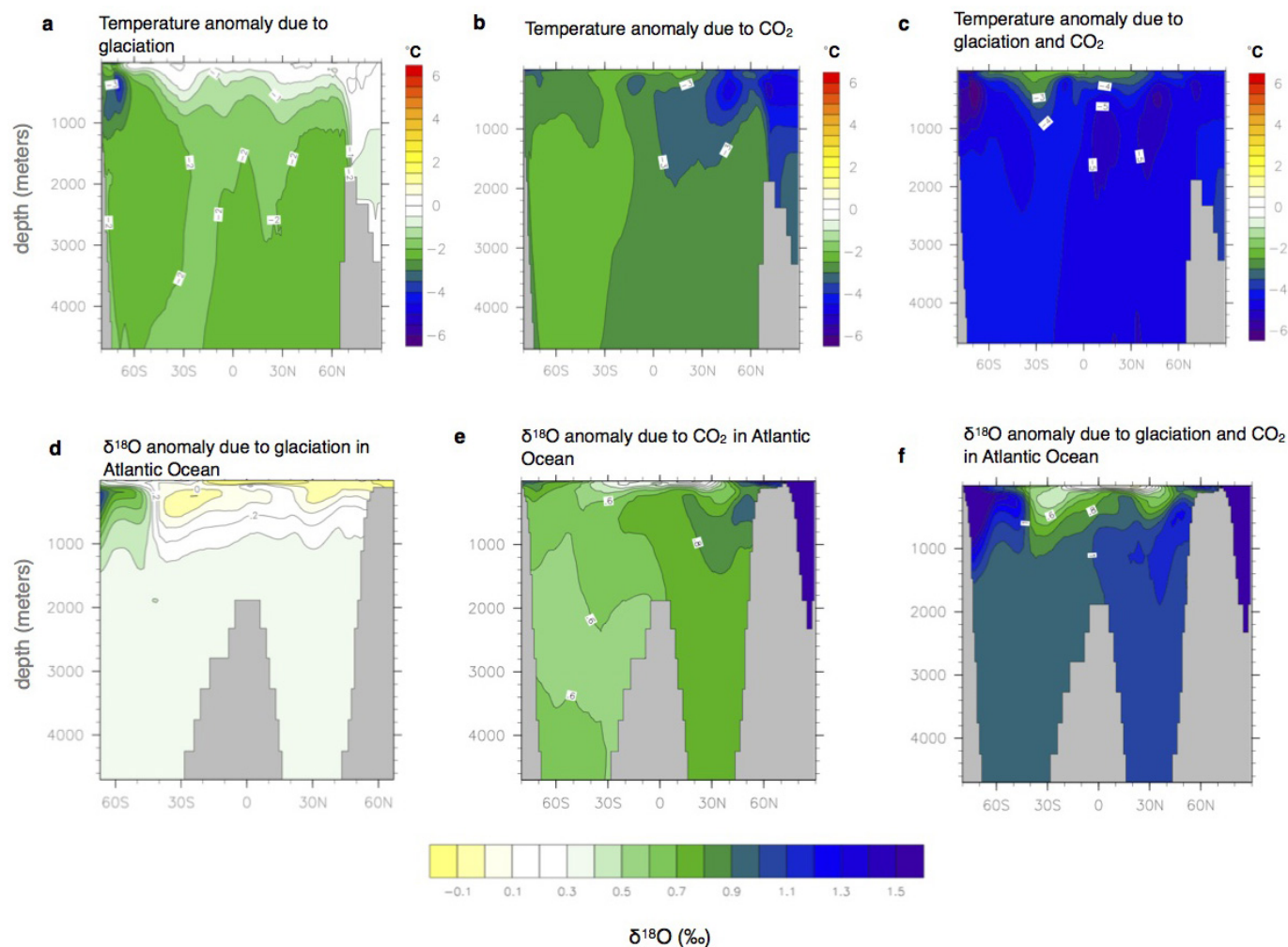
Recent work exploring the sensitivity of zonal mean OHT to changes in a wide range of alterations in boundary conditions, including ocean gateways, robustly shows small changes in OHT, similar to the results presented in this study⁴⁵.

$\delta^{18}\text{O}$ calculations. $\delta^{18}\text{O}$ values from the climate modelling were calculated throughout by using the temperature and salinity fields from the ocean model, using the empirical regression of ref. 46.

Carbon dioxide and ice-sheet forcing. Combining AIS and CO₂ forcing gives us our closest historical analogue to the EOT, and this change produces $\sim 1\%$ in $\delta^{18}\text{O}$ across the entire deep ocean basin (Extended Data Fig. 1). Results suggest that glaciation explains roughly 70% of the change in $\delta^{18}\text{O}$ across the southern high latitudes (Fig. 2c), whereas CO₂ explains the $\delta^{18}\text{O}$ change in the northern high latitudes (Extended Data Fig. 1).

The combination of both CO₂ and glaciation induces a more homogeneous cooling of both the surface and deep ocean (Extended Data Fig. 1c). This figure also shows the enhanced vertical $\delta^{18}\text{O}$ gradient exists (Extended Data Fig. 1f) even when both forcings (CO₂ and AIS) are combined, although the signal is reduced as a result of inclusion of the CO₂ forcing. This highlights the importance of separating the different forcings into pieces to fully understand their relative impact on temperature, $\delta^{18}\text{O}$ and ocean circulation. Recent work for the Middle Miocene transition has highlighted the important connections between ice sheet growth, changes in wind surface wind fields, and subsequent shifts in ocean circulation⁴⁷. The study was able to show that the growth of a large modern size ice sheet during the Middle Miocene is capable of producing deep ocean cooling while having a modest global surface temperature change.

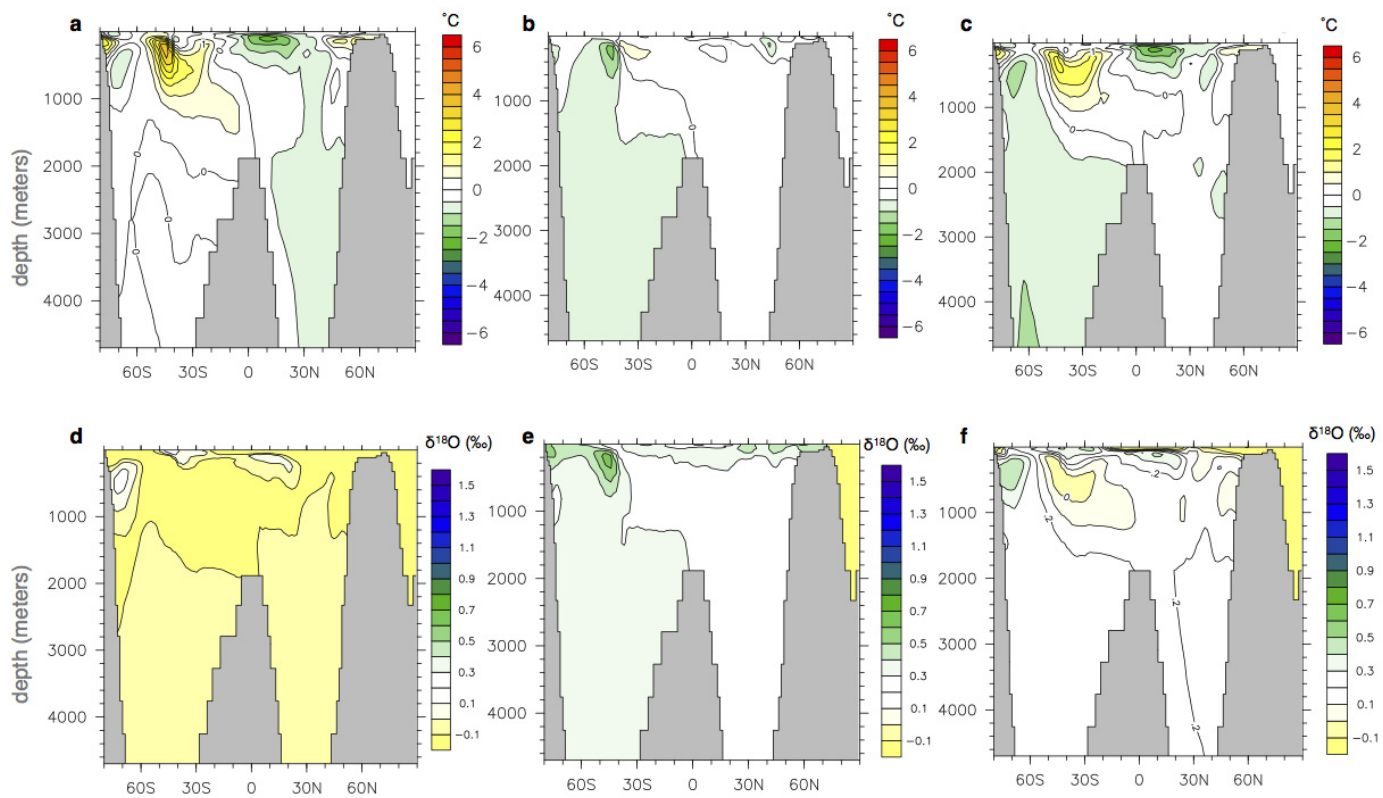
- Gent, P. R. *et al.* The Community Climate System Model version 4. *J. Clim.* **24**, 4973–4991 (2011).
- Lawrence, D. M. *et al.* The CCSM4 land simulation, 1850–2005: assessment of surface climate and new capabilities. *J. Clim.* **25**, 2240–2260 (2012).
- Bitz, C. M. *et al.* Climate sensitivity of the community climate system model version 4. *J. Clim.* **25**, 3053–3070 (2012).
- Large, W. G., Danabasoglu, G., McWilliams, J. C., Gent, P. R. & Bryan, F. O. Equatorial circulation in a global ocean climate model with anisotropic horizontal viscosity. *J. Phys. Oceanogr.* **31**, 518–536 (2001).
- Gent, P. R. & McWilliams, J. C. Isopycnal mixing in ocean circulation models. *J. Phys. Oceanogr.* **20**, 150–155 (1990).
- Jochum, M. Impact of latitudinal variations in vertical diffusivity on climate simulations. *J. Geophys. Res.* **114**, C01010 (2009).
- Ferrari, R. & Wunsch, C. Ocean circulation kinetic energy: reservoirs, sources, and sinks. *Annu. Rev. Fluid Mech.* **41**, 253–282 (2009).
- Danabasoglu, G., Ferrari, R. & McWilliams, J. C. Sensitivity of an ocean general circulation model to a parameterization of near-surface eddy fluxes. *J. Clim.* **21**, 1192–1208 (2008).
- Fox-Kemper, B., Ferrari, R. & Hallberg, R. Parameterization of mixed layer eddies. I. Theory and diagnosis. *J. Phys. Oceanogr.* **38**, 1145–1165 (2008).
- Large, W. G. & Danabasoglu, G. Attribution and impacts of upper-ocean biases in CCSM3. *J. Clim.* **19**, 2325–2346 (2006).
- Ali, J. R. & Huber, M. Mammalian biodiversity on Madagascar controlled by ocean currents. *Nature* **463**, 653–656 (2010).
- Heavens, N. G., Shields, C. A. & Mahowald, N. M. A paleogeographic approach to aerosol prescription in simulations of deep time climate. *J. Adv. Model. Earth Syst.* **4**, M11002 (2012).
- Wilson, D. S., Pollard, D., DeConto, R. M., Jamieson, S. S. R. & Luyendyk, B. P. Initiation of the West Antarctic Ice Sheet and estimates of total Antarctic ice volume in the earliest Oligocene. *Geophys. Res. Lett.* **40**, 4305–4309 (2013).
- Eldrett, J. S., Greenwood, D. R., Harding, I. C. & Huber, M. Increased seasonality through the Eocene to Oligocene transition in northern high latitudes. *Nature* **459**, 969–973 (2009).
- Hill, D. J. *et al.* Paleogeographic controls on the onset of the Antarctic circumpolar current. *Geophys. Res. Lett.* **40**, 5199–5204 (2013).
- Broecker, W. S. The salinity contrast between the Atlantic and Pacific oceans during glacial time. *Paleoceanography* **4**, 207–212 (1989).
- Knorr, G. & Lohmann, G. Climate warming during Antarctic ice sheet expansion at the Middle Miocene transition. *Nature Geosci.* **7**, 376–381 (2014).



Extended Data Figure 1 | Ocean temperature and δ¹⁸O anomalies.

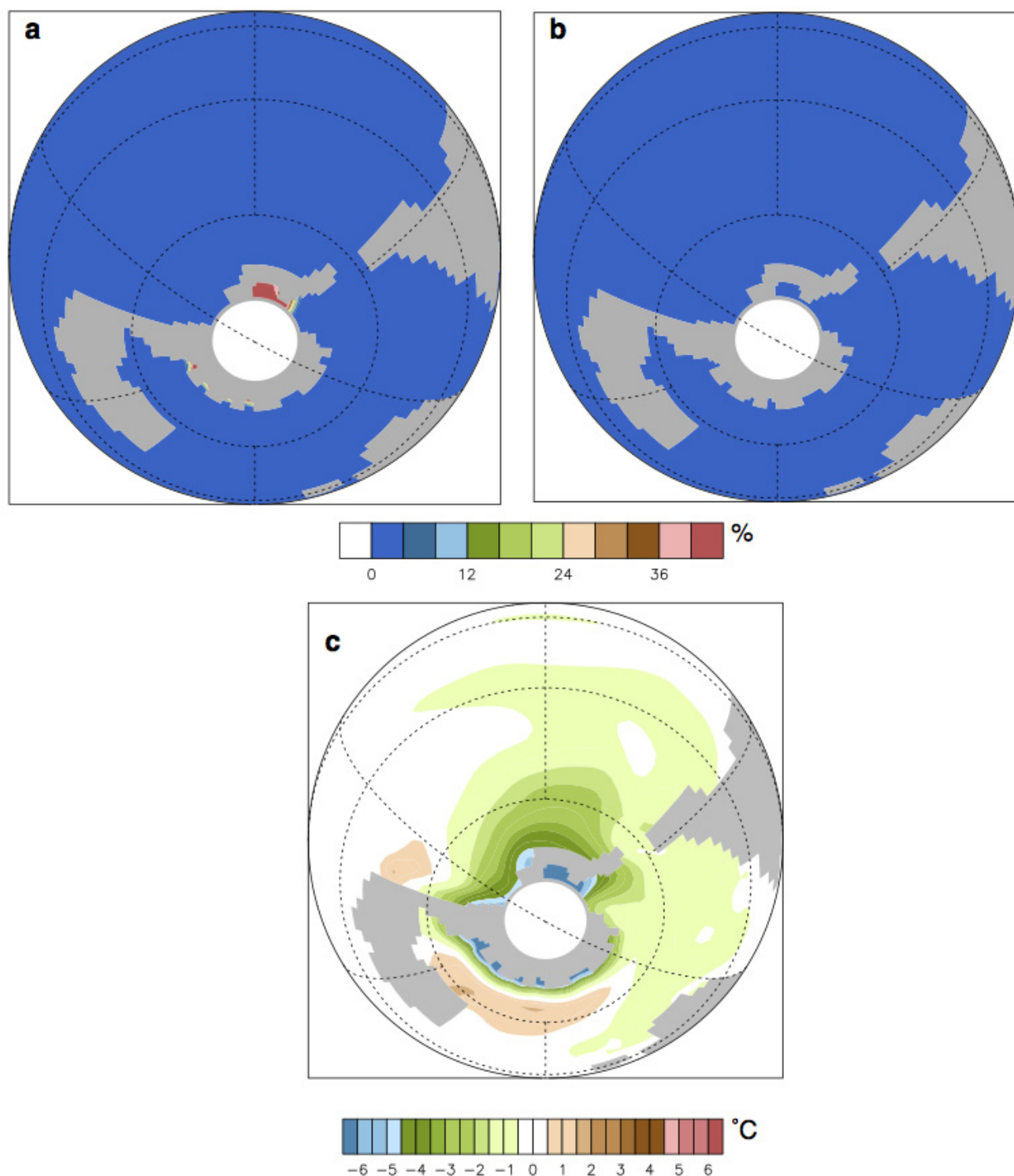
a–c, Zonally averaged temperature anomalies (°C) averaged over all longitudes: Glaciated minus unglaciated cases (CO₂ constant at 560 p.p.m.) (**a**), 1,120 minus 560 p.p.m. CO₂ cases (both unglaciated) (**b**), and unglaciated case with

1,120 p.p.m. CO₂ minus glaciated case with 560 p.p.m. CO₂ (**c**). **d–f**, δ¹⁸O comparisons (per mil) zonally averaged over the Atlantic basin; otherwise as in **a–c**.



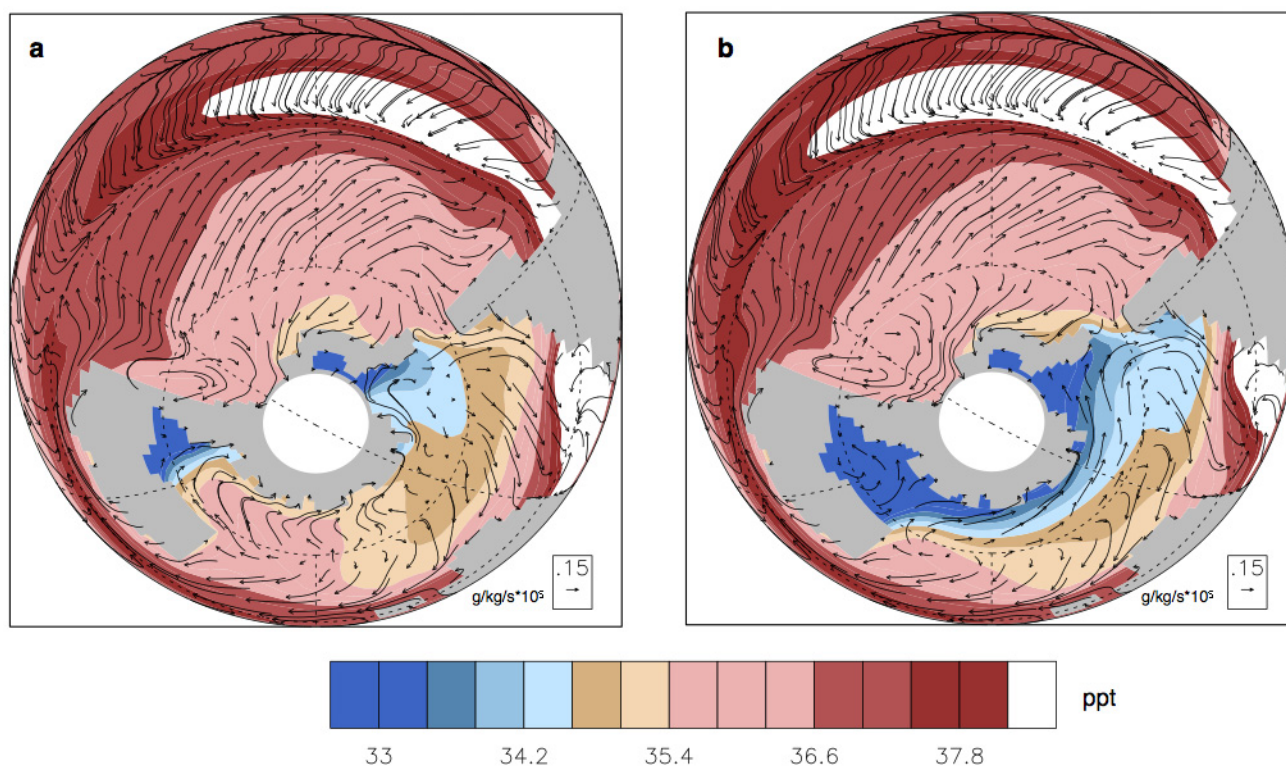
Extended Data Figure 2 | Ocean temperature and $\delta^{18}\text{O}$ anomalies in Atlantic Ocean basin due to Southern Ocean gateway opening.
a, Temperature anomaly (°C) for both gateways opened minus both gateways

closed. **b**, Temperature anomaly for DP and TG closed minus DP open and TG closed. **c**, Temperature anomaly for DP and TG open minus TG closed DP open. **d-f**, As in a-c, except for $\delta^{18}\text{O}$.

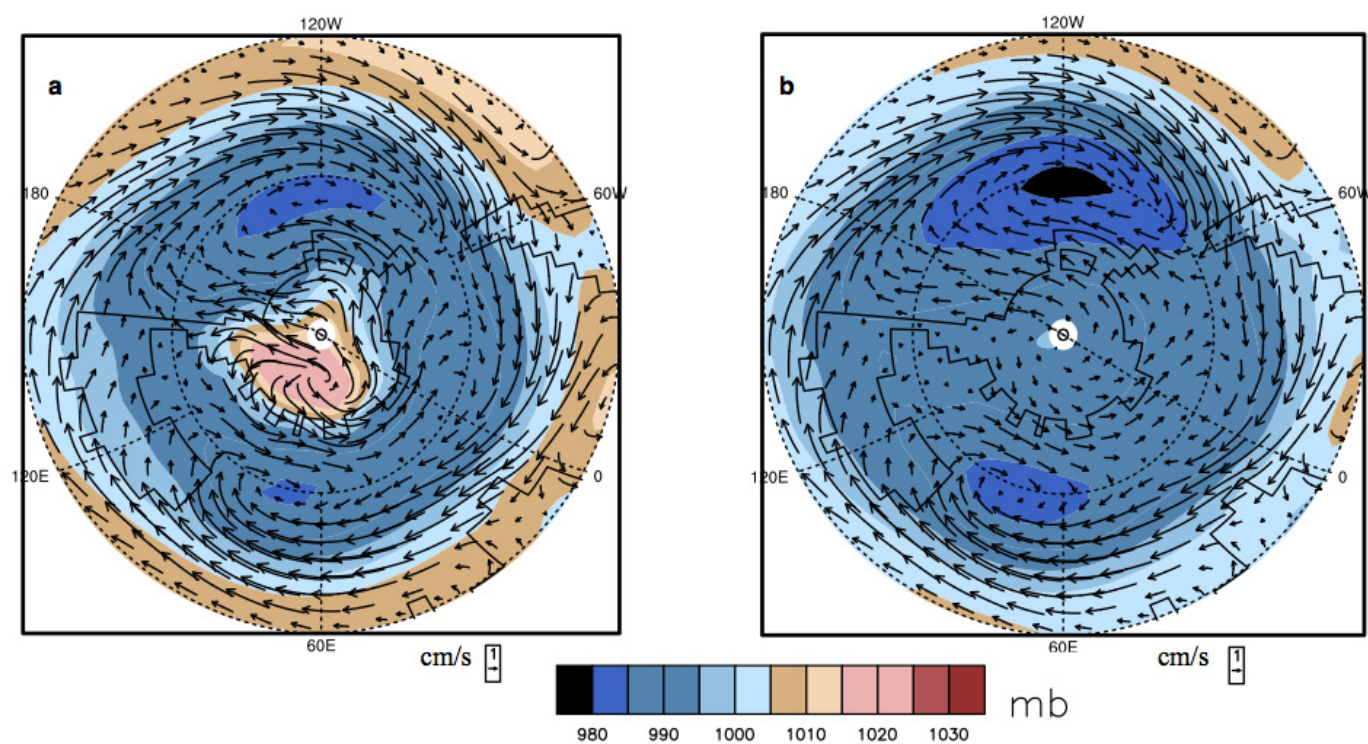


Extended Data Figure 3 | Absolute sea-ice and anomalous sea surface temperature. a, b, Sea ice fraction for glaciased (a) and unglaciased (b) cases.

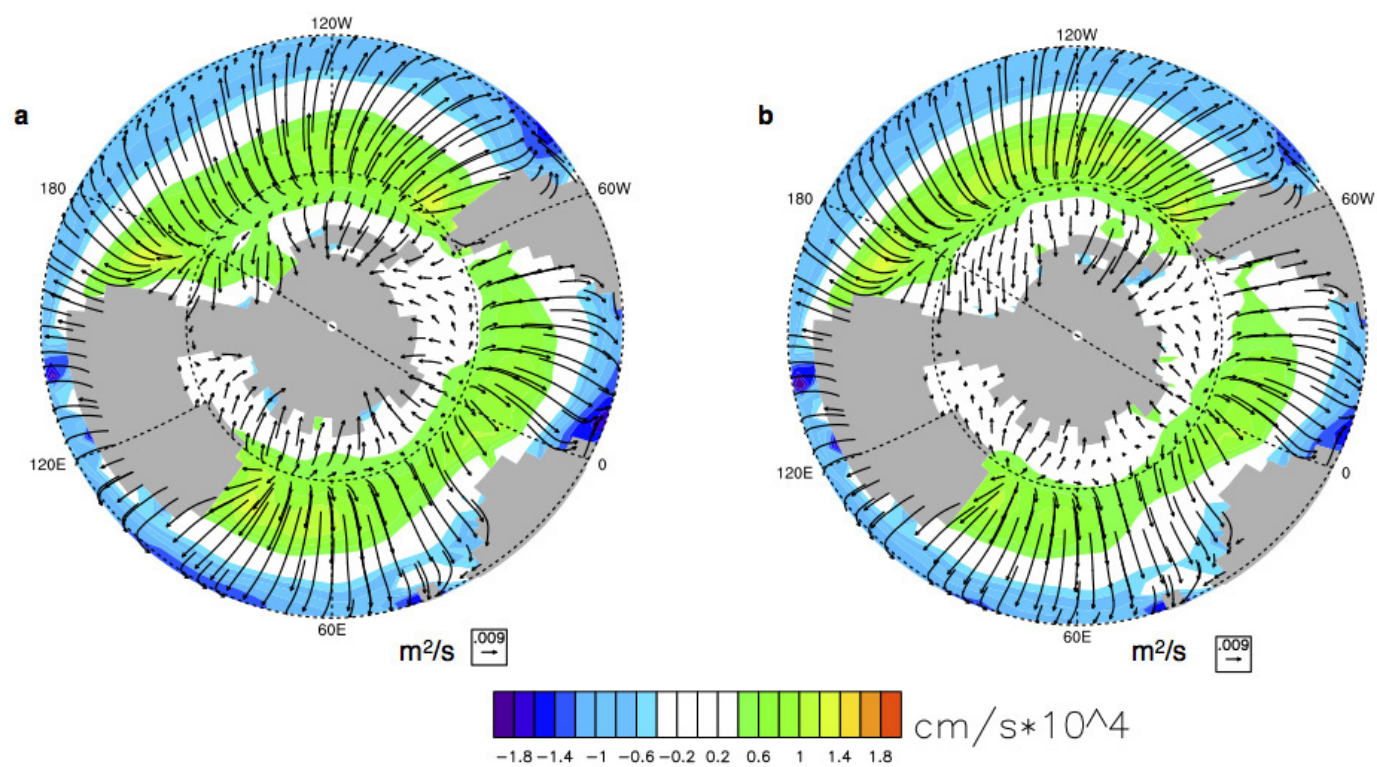
c, Glaciased minus unglaciased sea surface temperature anomaly. All simulations with 1,120 p.p.m. CO₂.



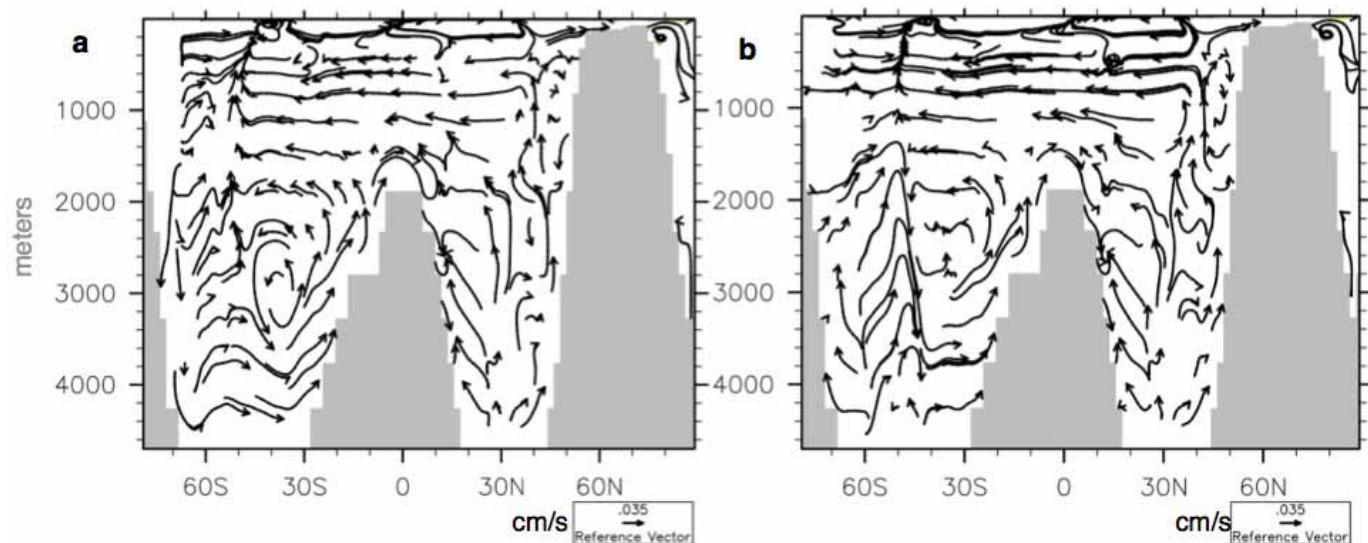
Extended Data Figure 4 | Absolute salinity fields. Salinity (colour contour) and salinity flux (vectors) for glaciated late Eocene (a) and unglaciated cases (b) at $1,120 \text{ p.p.m. CO}_2$.



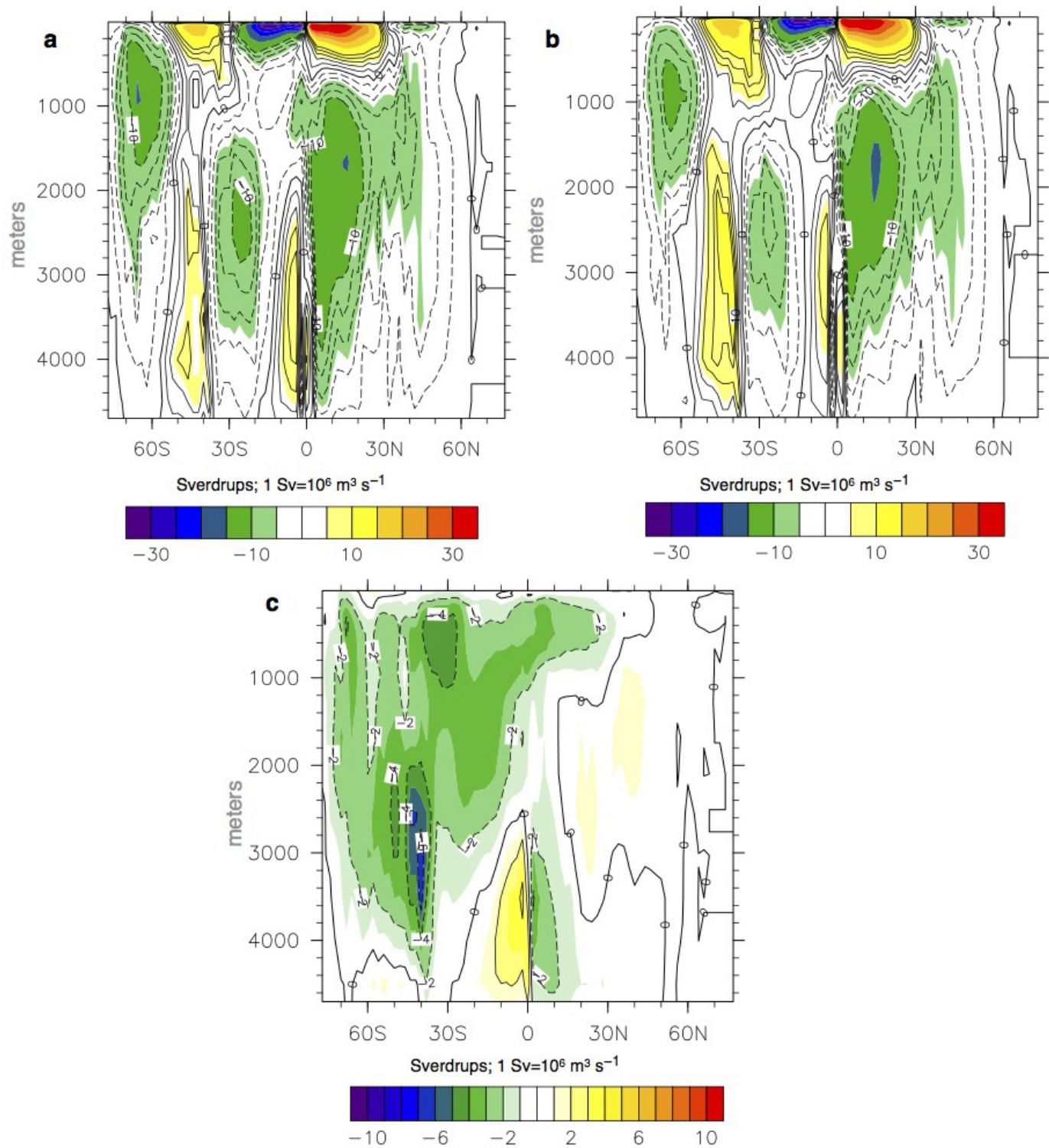
Extended Data Figure 5 | Absolute sea level pressure and surface wind. Sea level pressure (colour contour) and surface wind (vectors) for glaciased (a) and unglaciased (b) cases at 1,120 p.p.m. CO_2 .



Extended Data Figure 6 | Absolute Ekman pumping and transport. Ekman pumping contour and Ekman transport overlaid as vectors for glaciated (a) and unglaciated (b) cases at 1,120 p.p.m. CO₂. See the calculations for Ekman pumping and transport in Methods.

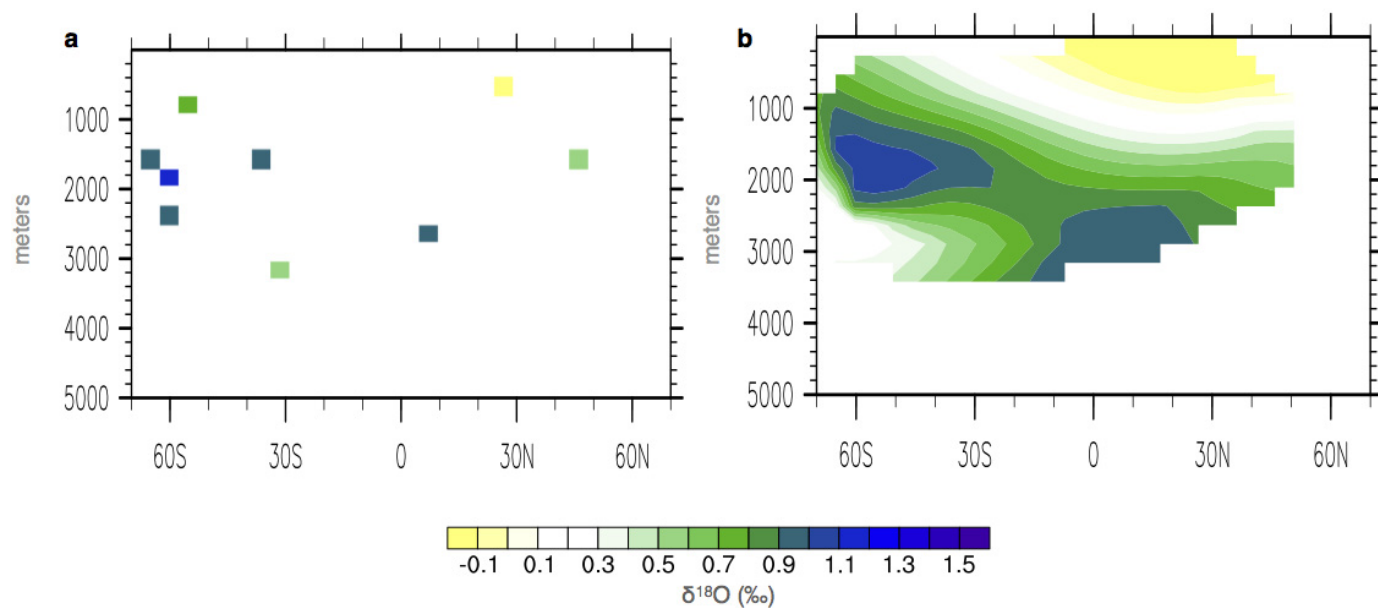


Extended Data Figure 7 | Absolute ocean currents. Zonally averaged ocean currents (meridional and vertical) across the Atlantic Ocean. The vertical ocean velocities are scaled by a constant coefficient (500) for plotting purposes. Glaciated (a) and unglaciated (b) cases at 1,120 p.p.m. CO₂.



Extended Data Figure 8 | Meridional overturning circulation. Zonally averaged meridional overturning circulation anomaly for glaciated

(a) unglaciated, (b) and glaciated minus unglaciated (c) case anomaly at 1,120 p.p.m. CO_2 .



Extended Data Figure 9 | Depth-latitude plot for the non-interpolated and interpolated $\delta^{18}\text{O}$ proxy record anomalies. **a**, Raw $\delta^{18}\text{O}$ anomalies. **b**, Interpolated $\delta^{18}\text{O}$ anomalies (see Extended Data Table 1).

Extended Data Table 1 | Site locality name, palaeolatitude, palaeolongitude, palaeodepth and $\delta^{18}\text{O}$

Site Number	Paleo-Lat	Paleo-Lon	Depth	$\delta^{18}\text{O}$
DSDP366	5.74	-23.17	2710	0.99
DSDP522	-31.69	-3.82	3039	0.58
DSDP529	-34.29	-1.87	1672	0.97
DSDP549	45.90	-13.10	1659	0.59
ODP689	-65.11	2.88	1650	0.99
ODP744	-60.41	78.54	2313	0.97
ODP690	-66.24	1.75	2531	0.18
ODP748	-57.42	77.72	800	0.74
1053A	28.10	-69.90	400	-0.28
ODP738	-61.05	81.53	1750	1.1

Proxy records compiled from refs 2, 4. Positive values indicate an increase in $\delta^{18}\text{O}$ going from the late Eocene (34–36 Myr ago) into the early Oligocene (31–29 Myr ago).

Widespread mixing and burial of Earth's Hadean crust by asteroid impacts

S. Marchi¹, W. F. Bottke¹, L. T. Elkins-Tanton^{2†}, M. Bierhaus³, K. Wünnemann³, A. Morbidelli⁴ & D. A. Kring⁵

The history of the Hadean Earth (~4.0–4.5 billion years ago) is poorly understood because few known rocks are older than ~3.8 billion years old¹. The main constraints from this era come from ancient submillimetre zircon grains^{2,3}. Some of these zircons date back to ~4.4 billion years ago when the Moon, and presumably the Earth, was being pummelled by an enormous flux of extraterrestrial bodies⁴. The magnitude and exact timing of these early terrestrial impacts, and their effects on crustal growth and evolution, are unknown. Here we provide a new bombardment model of the Hadean Earth that has been calibrated using existing lunar⁴ and terrestrial data⁵. We find that the surface of the Hadean Earth was widely reprocessed by impacts through mixing and burial by impact-generated melt. This model may explain the age distribution of Hadean zircons and the absence of early terrestrial rocks. Existing oceans would have repeatedly boiled away into steam atmospheres as a result of large collisions as late as about 4 billion years ago.

Terrestrial planet formation models indicate the Earth went through a sequence of major growth phases: accretion of planetesimals and planetary embryos over many tens of millions of years (see, for example, ref. 6), culminating in a final giant impact that led to the formation of our Moon (see, for example, ref. 7). This was followed by the late accretion of left-over planetesimals that probably contributed less than 0.5% of the Earth's present-day mass⁵. Although the role of late accretion impacts on the Hadean Earth has long been discussed (for example, in ref. 8), the precise nature of the impactor flux during late accretion is elusive.

Estimates from the abundance of highly siderophile elements (HSEs, such as Re, Au, Os and Ru) in mantle-derived peridotites indicate that $\sim(0.7\text{--}3.0) \times 10^{22}$ kg of material with broad chondritic composition was added to the Earth⁵, probably during the late accretion phase (see Methods).

An additional constraint on this flux comes from the ratio of HSEs found in the mantles of the Earth and Moon. Studies of terrestrial and lunar samples suggest that the ratio of the mass of broadly chondritic material accreted by the Earth and the Moon is probably $\geq 700:1$ (refs 9, 10). By modelling the impactor flux on both worlds, it has been argued¹⁰ that this ratio was a reasonable outcome of stochastic accretion, with most HSEs added to the Earth by massive impactors that were statistically unlikely to strike the smaller Moon. This scenario was recently found to be broadly consistent with the current generation of models of terrestrial planet formation¹¹.

Here we assess the early Earth's impact history by rescaling a recent estimate of the lunar impact flux⁴ to Earth. The advantages of this approach are numerous. First, the Moon provides a much clearer record of the early impact history of the Earth–Moon system⁸. Moreover, the lunar cratering record provides an absolute impactor flux that is independent of assumptions made by terrestrial planet formation models. The rescaling was done transforming lunar craters into a projectile flux, with the flux used to estimate the number of terrestrial impactors taking place in intervals of 25 Myr between 3.5 and 4.5 Gyr ago (see Methods). For the purpose of our work we assume that the Moon-forming impact was at

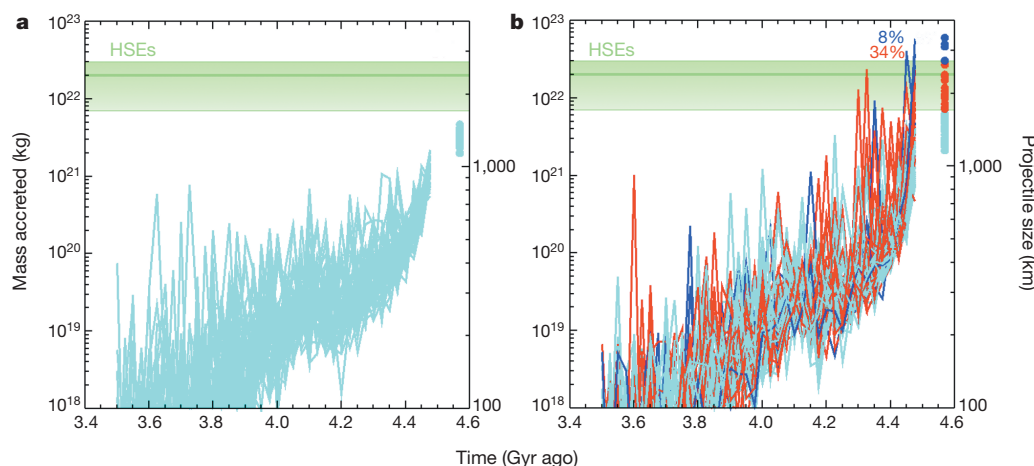


Figure 1 | Mass accreted by the Earth during the late accretion phase.

a, Cyan curves show 50 representative Monte Carlo simulations corresponding to the main-belt size–frequency distribution truncated at Ceres (see Extended Data Fig. 1). Each data point indicates the total mass (left y axis) and equivalent diameter (right y axis) accreted in that time bin (of 25 Myr each). The cumulative accreted mass (or equivalent size) in the period 3.5–4.5 Gyr ago is indicated by the dots at the right of the panel. We assumed a projectile density

of $3,000 \text{ kg m}^{-3}$. The horizontal green lines mark the lower, most probable and upper limits for the accreted mass as inferred from HSEs⁵. **b**, As in **a**, but for an impactor size–frequency distribution extrapolated at 4,000 km before 4.15 Gyr ago (see Extended Data Fig. 1). Simulations that deliver a mass within the HSEs range are in red; those in excess of the maximum limit are in blue (the corresponding percentages of simulations are indicated). These simulations were retained for further analyses.

¹Southwest Research Institute, Boulder, Colorado 80302, USA. ²Carnegie Institution for Science, Washington DC 20015, USA. ³Museum für Naturkunde, Berlin 10115, Germany. ⁴Observatoire de la Côte d'Azur, Nice 06304, France. ⁵Universities Space Research Association, Lunar and Planetary Institute, Houston, Texas 77058, USA. [†]Present address: School of Earth and Space Exploration, Arizona State University, Tempe, Arizona 85287, USA.

~ 4.5 Gyr ago, but our results are insensitive to the exact timing. The sizes of the projectiles were randomly drawn from an assumed impactor size–frequency distribution (SFD) with a shape similar to that of large main-belt asteroids. The cut-off of this population was varied for different time intervals. Impactors striking after ~ 4.15 Gyr ago, the putative starting time of the late heavy bombardment (LHB^{4,12–14}) were given a maximum cut-off of 1,000 km, roughly corresponding to the largest present-day asteroid Ceres. The impactor SFD before ~ 4.15 Gyr ago is assumed to have larger left-over planetesimals. We assumed a similar impactor SFD with a cut-off threshold at 4,000 km, which was extrapolated from the SFD of inner and central main-belt asteroids ranging between a few hundred kilometres to 1,000 km (see Methods and Extended Data Figs 1 and 2).

Using a Monte Carlo code, we repeated this procedure $\sim 5,000$ times to address the stochastic variability intrinsic to late accretion projectiles, and computed the accreted mass (Fig. 1). A key result is that, for the case of impactor SFD cut-off at 1,000 km, the total delivered mass is

always below the range expected from HSEs (Fig. 1a), whereas a significant fraction of simulations ($\sim 30\%$) fall in this range when the cut-off is at 4,000 km before 4.15 Gyr ago (Fig. 1b). We assumed that all projectiles and their HSEs were fully accreted into the silicate Earth. Although the latter assumption is probably true for small projectiles that disintegrate in the impact event, larger objects may not deliver their HSEs efficiently to the mantle in all circumstances. Assuming that up to 50% of the cores of planetesimals $\geq 2,000$ km may be lost (see Methods), we compute that the corresponding total percentage of successful simulations may be as high as $\sim 40\%$. These results elucidate several important features of late accretion. The bulk of the mass (99% and 90%) is delivered by $\sim 17 \pm 5$ and $\sim 6 \pm 3$ largest projectiles, respectively, and the largest impactors can exceed $\sim 3,000$ km. This is therefore a highly stochastic regime, in which a few projectiles dominate the budget of HSEs delivered to Earth, in agreement with previous findings¹⁰. Moreover, by tracking the timing of the impacts, we also find that most of the mass is typically accreted over a significant fraction of the Hadean (~ 4.2 – 4.5 Gyr

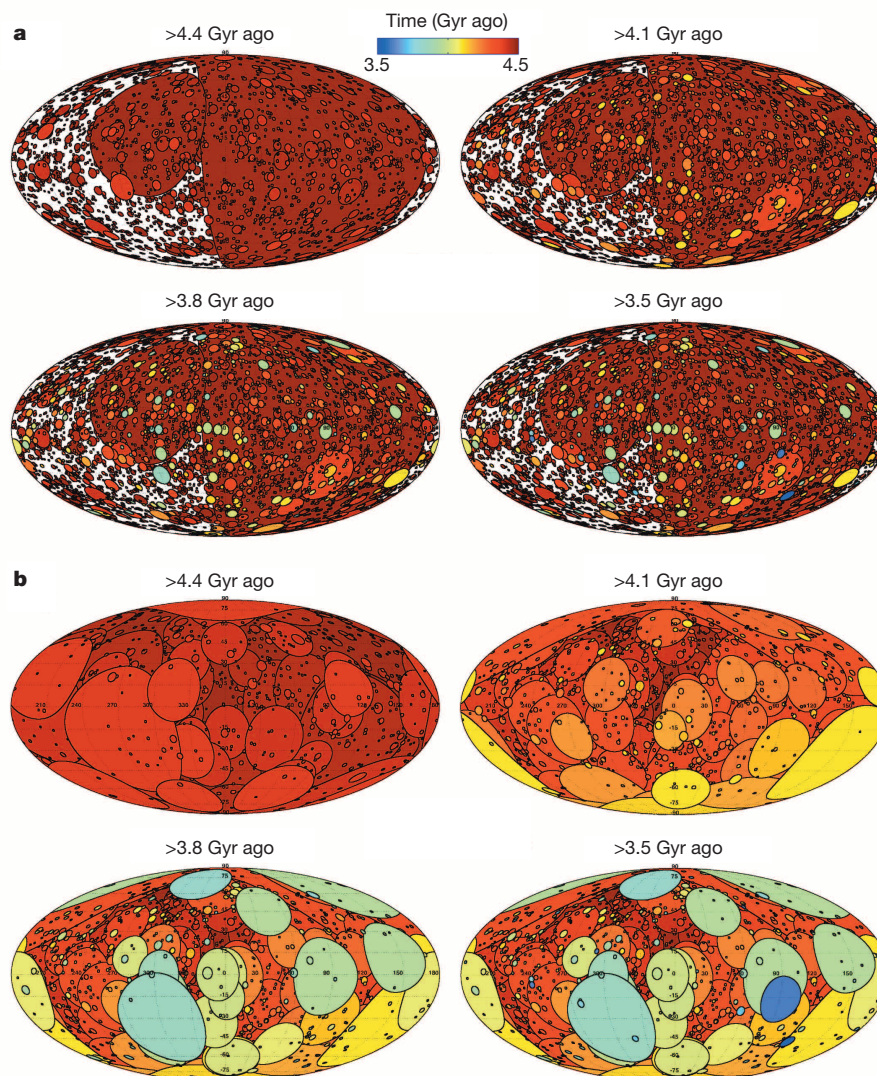


Figure 2 | Spatial distribution and sizes of craters formed on the early Earth. **a**, Mollweide projections of the cumulative record of craters at four different times. Each circle indicates the final crater size estimated from the transient cavity size from our simulations and a conservative estimate for the transient-to-final crater size scaling (see Methods). The maps do not show ejecta blankets and melt extrusion on the surface, which can greatly expand the effects of

cratering; they also do not account for a hotter early geotherm, which would also result in larger crater sizes (see Methods). The colour coding indicates the time of impact. The smallest projectiles considered have a diameter of 15 km. We assumed an impact velocity of ~ 16 and $\sim 25 \text{ km s}^{-1}$ before and after 4.15 Gyr ago (see Methods), respectively, and a most likely impact angle of 45° . **b**, As in **a**, but including melt extrusion on the surface as discussed in the text.

ago), suggesting that the early Hadean silicate Earth could have had a substantially different budget of HSEs and trace elements compared with the current Earth (see Methods). Note also that less than 0.5% of the simulations deliver more than 1% of an Earth mass.

Our terrestrial bombardment model also sheds light on the role of impacts on the geological evolution of the Hadean Earth, with particular emphasis on mixing, burial and melting of the uppermost layers. To model these effects quantitatively, we performed a suite of impact simulations with the Simplified Arbitrary Lagrangian Eulerian shock-physics code (iSALE¹⁵), and computed the resulting excavation cavity size, excavation volume and depth, and the volume of target melt. We varied the target temperatures and considered impactor diameters ranging from 15 to 4,000 km with a range of impact velocities (see Methods).

A key process is the impact-generated mixing as a result of the excavation and collapse of large transient cavities in the lithosphere. We found that before ~ 4.4 Gyr ago up to 60–70% of the Earth's surface was reworked to a median depth of 20 km (Fig. 2a). Thus, our model predicts prolonged crustal reworking and mixing of various components, as inferred from recent Pb–Hf isotope systematics of Hadean zircons¹⁶.

Melting of the target is an additional important process. We computed the melt produced by shock pressure, using both analytical estimates^{17,18} and iSALE simulations. Our simulations agree well with analytical estimates for impactor sizes below 100 km, but deviate significantly for larger impactors (Fig. 3a) (see Methods and Extended Data Figs 3 and 4). Both methods neglect impact-induced decompression and subsequent adiabatic melting of rising material in the mantle, which increase the total volume of melt. These processes have been quantitatively modelled for impactor diameters smaller than 100 km (ref. 19), and here we use their predictions for an impactor 100 km in diameter and then extrapolate their results to larger projectiles (Fig. 3a). Note that the melt volume is a lower limit for large projectiles that are expected to induce major mantle perturbations, resulting in voluminous adiabatic melting²⁰. Computed melt volumes greatly exceed the volumes of current flood basaltic provinces¹⁹. Some fraction of the mantle melt will erupt; through isostatic adjustment, melt may be expelled from the shallowing crater onto the planetary crust¹⁷. Melt spreading is also aided by the dynamics of cavity collapse in a hotter crust²¹, such as that envisioned to have occurred during the Hadean²². Assuming that the impact-generated melt

flows on the surface, we used the estimated melt volumes to calculate that the corresponding diameter of a spherical cap with a thickness of 3 km (comparable to large terrestrial igneous provinces) is ~ 20 – 30 -fold that of the impactor diameters (see Fig. 3b). As a result of melt spreading, lithologies previously exposed at the surface are buried over large areas.

The effects of melt burial due to impacts are shown in Fig. 2b. The cumulative fraction of Earth's surface buried by impact-generated melt is 70–100% since 4.15 Gyr ago, and it increases to 400–600% during the period 4.15–4.5 Gyr ago. These findings do not preclude the possibility of having large unaffected surface areas at any given time step (see Fig. 2b, Extended Data Fig. 5), a condition required for liquid water to be stable at the surface as indicated by $\delta^{18}\text{O}$ measurements in Hadean zircons^{23,24}.

Additional constraints on the terrestrial bombardment flux may come from trace elements entrapped in Hadean zircons. Their rare-earth elements, U–Pb and Pb–Hf ages and Lu/Hf ratios point to significant mixing of mafic and felsic reservoirs (see, for example, ref. 16). This mixing is sometimes attributed to volcanism or subduction^{3,16,22}, in which weathered upper crustal reservoirs are buried at depth. Once buried, the different components melt, and the resulting magma can then crystallize Hadean zircons³. It is unclear, however, whether these processes can readily explain the observed age distribution of Hadean zircons, which is characterized by a well-defined peak at 4.1–4.2 Gyr ago and a lack of ages older than ~ 4.4 Gyr (ref. 2). Moreover, it is unresolved whether the lack of ultra-ancient zircons implies that the right conditions for zircon formation were not met during this time, or whether zircons older than ~ 4.4 Gyr did not survive subsequent evolution.

Our model shows that substantial burial could be achieved by impact-generated melt. Assuming that burial is required to make Hadean zircons^{3,16,22}, we investigated whether the burial by impact-generated melt could explain the Hadean zircons age distribution. In our Monte Carlo code, each simulated impact was assumed to bury surface lithologies within a threshold distance proportional to the projectile diameter d multiplied by a factor f ($\sim fd$). This process results in increased crustal temperatures over a large annulus around the impact site, possibly leading to eutectic melting of buried wet crustal material, in agreement with the observation that many of the Hadean zircons probably crystallized

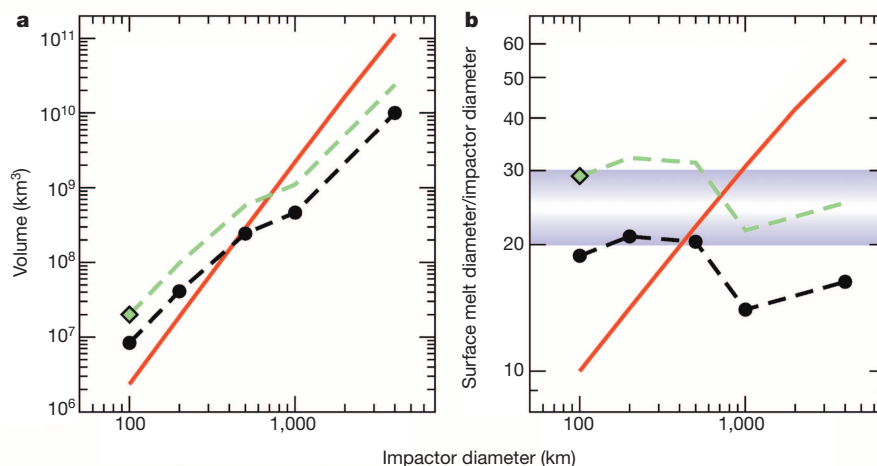


Figure 3 | Melt production by large impacts on the Earth. **a**, Impact-shock melt volume from analytical estimates¹⁸ (red line), and impact-generated melt volume (including shock and decompression melting) for impacts simulated with iSALE (black dots). These simulations assumed a planar target, a lithospheric thickness of 125 km, a mantle potential temperature of 1,400 °C, an impact velocity of 12.7 km s^{−1} (corresponding to 18 km s^{−1} for an impact angle of 45°). The mantle potential temperature of the Hadean Earth may have been hotter than assumed here by ~ 200 °C (see Methods and Extended Data Table 1). We also ran simulations for a mantle potential temperature of 1,600 °C and found that the melt volume increased by 75% (see Methods). The

green diamond represents the total melt volume (including decompression and adiabatic melting) from ref. 19 for a crater 800 km in diameter (corresponding to an impactor of 100 km for the assumed impact conditions) and a mantle potential temperature of 1,450 °C. The latter data point is 2.3-fold higher than our data point. Assuming that a similar scaling holds for larger projectiles, we obtain the green curve. The curves neglect adiabatic melting and therefore provide a lower limit for the total impact-generated melt. **b**, Ratio of surface melt diameter (for a thickness of 3 km) to impactor diameter (f). The horizontal grey area indicates $f = 20$ – 30 , as discussed in the text. Symbols as in **a**.

from wet eutectic melts³. The increase in crustal temperatures is caused by the hotter geotherm of the buried material produced by melt above and possible thinning from the bottom, the rise of mantle melts into the fractured and tectonized lithosphere close to the crater's rims^{19,25}, and possibly also by dripping crustal diapirs originating from the thick surface melt layer at larger radial distances²⁶.

We find that for $f \approx 20$ – 30 for impactor diameters larger than 100 km, as predicted by our simulations, the resulting surface age distribution matches well the Hadean zircon age distribution (see Fig. 4). In contrast, simulations having $f \lesssim 20$ or $f \gtrsim 40$ fail to reproduce the data (see Methods). This fit, if not a coincidence, tells us that large projectiles were capable of making zircons and resetting pre-existing ones over regions well beyond their computed crater rims, in agreement with our estimates of impact-generated melt (Fig. 3).

Several observations may be explained by consequences of the proposed mechanism for Hadean zircons formation. First, given that large impactors could have struck at relatively late times (for example the expected surge of projectiles at 4.15 Gyr ago via the LHB), zircon production through these impact-generated processes could have occurred for many hundreds of millions of years, as observed^{2,27}. Attrition among the oldest zircons was pronounced, because they were subject to high temperatures near numerous impact locations²⁸, burial at depth by melt and ejecta, and redistribution to the upper crust. Most were reset or destroyed. Consequently, our model predicts that the paucity of zircons older than 4.4 Gyr is expected from collisional processes. Moreover, the formation of Hadean zircons at depth in large annuli around major impacts provides a ready explanation for their lack of clear signs of impact shock^{2,27} that are commonly observed among younger zircons. Another attractive aspect of our model is that it explains the mixing of the protoliths from which Hadean zircons crystallized, as inferred from the Hf-isotope record²⁹. Finally, the volume of the buried lithologies at any time step is about an order of magnitude higher than the fraction of the crust melted by small impacts, indicating that the zircon crystallization from impact melts was negligible (see Methods), in agreement with the low crystallization temperatures observed in Hadean zircon^{30,31}.

We argue that the peak of Hadean zircon ages at 4.1–4.2 Gyr reflects the onset of the LHB, as suggested by meteorite Ar–Ar shock degassing ages and other data^{13,14}. Indeed, we find that a scenario with an LHB spike at significant younger ages, say 3.9 Gyr, is inconsistent with Hadean zircon

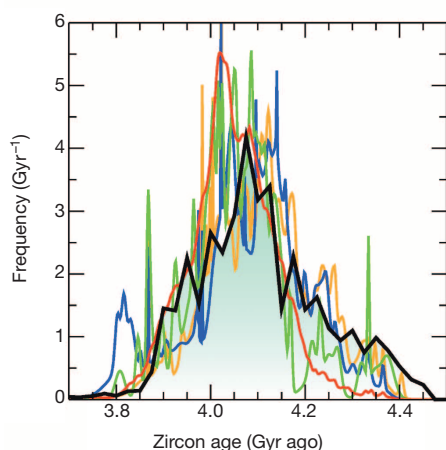


Figure 4 | Detrital Hadean zircon ages compared with the computed distribution of impact-generated ages. Zircon ages (coloured curves correspond to different data sets: orange, ^{207}Pb – ^{206}Pb ages²; blue, U–Pb ages²⁹; green, ^{207}Pb – ^{206}Pb ages¹⁶; red, U–Pb ages³²) show a distinct peak at ~ 4.1 – 4.2 Gyr ago. In agreement with our iSALE simulations, the distribution of impact-generated ages (black line, shaded area) is computed for $f = 30$ for projectiles larger than 100 km and $f = 9$ for projectiles smaller than 100 km, and is an average of 50 successful Monte Carlo simulations (Fig. 1). All distributions are normalized to unit area.

age distributions. A similar conclusion is reached for a steadily declining bombardment (no LHB) scaled to match the abundance of lunar HSEs (see Methods). Therefore, LHB-era impactors provide a natural explanation for the clustering of Hadean zircon ages that would otherwise require *ad hoc* endogenic conditions (for example increased subduction or volcanism rates).

The new picture of the Hadean Earth emerging from our work has important implications for its habitability. Before ~ 4 Gyr ago, no substantial large region of the Earth's surface could have survived untouched by impacts and associated outcomes. Large impacts had particularly severe effects on extant ecosystems. We find that the Hadean was plausibly characterized by one to four impactors larger than 1,000 km capable of global sterilization⁸, and by three to seven impactors larger than 500 km capable of global ocean vaporization⁸. The median time for the latest impactor larger than 500 km to hit the Earth was ~ 4.3 Gyr ago. In $\sim 10\%$ of the simulations, this could be as recently as ~ 4 Gyr ago (Extended Data Fig. 6), depending on various assumptions. Thus, life emerging during the Hadean was probably resistant to high temperatures and was capable of spreading from the stable niches that existed at that time.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 4 March; accepted 23 May 2014.

- Kamber, B. S., Moorbath, S. & Whitehouse, M. J. The oldest rocks on Earth: time constraints and geological controversies. *Geol. Soc. Lond. Spec. Publ.* **190**, 177–203 (2001).
- Cavosie, A. J., Valley, J. W. & Wilde, S. A. In *World's Oldest Rocks* (ed. Condie, K.) 111–129 (Developments in Precambrian Geology vol. 15, Elsevier, 2007).
- Harrison, T. M. The Hadean Crust: evidence from >4 Ga zircons. *Annu. Rev. Earth Planet. Sci.* **37**, 479–505 (2009).
- Morbidelli, A., Marchi, S., Bottke, W. F. & Kring, D. A. A sawtooth-like timeline for the first billion years of lunar bombardment. *Earth Planet. Sci. Lett.* **355–356**, 144–151 (2012).
- Walker, R. J. Highly siderophile elements in the Earth, Moon and Mars: update and implications for planetary accretion and differentiation. *Chem. Erde* **69**, 101–125 (2009).
- Chambers, J. E. Planetary accretion in the inner Solar System. *Earth Planet. Sci. Lett.* **223**, 241–252 (2004).
- Canup, R. M. Dynamics of lunar formation. *Annu. Rev. Astron. Astrophys.* **42**, 441–475 (2004).
- Sleep, N. H., Zahnle, K. J., Kasting, J. F. & Morowitz, H. J. Annihilation of ecosystems by large asteroid impacts on the early earth. *Nature* **342**, 139–142 (1989).
- Day, J. M. D., Pearson, D. G. & Taylor, L. A. Highly siderophile element constraints on accretion and differentiation of the Earth–Moon system. *Science* **315**, 217–219 (2007).
- Bottke, W. F., Walker, R. J., Day, J. M. D., Nesvorný, D. & Elkins-Tanton, L. Stochastic late accretion to Earth, the Moon, and Mars. *Science* **330**, 1527–1530 (2010).
- Raymond, S. N., Schlichting, H. E., Hersant, F. & Selsis, F. Dynamical and collisional constraints on a stochastic late veneer on the terrestrial planets. *Icarus* **226**, 671–681 (2013).
- Marchi, S. *et al.* Global resurfacing of Mercury 4.0–4.1 billion years ago by heavy bombardment and volcanism. *Nature* **499**, 59–61 (2013a).
- Bottke, W. F. *et al.* An Archaean heavy bombardment from a destabilized extension of the asteroid belt. *Nature* **485**, 78–81 (2012).
- Marchi, S. *et al.* High-velocity collisions from the lunar cataclysm recorded in asteroidal meteorites. *Nature Geosci.* **6**, 303–307 (2013b).
- Wünnemann, K., Collins, G. S. & Melosh, H. J. A strain-based porosity model for the use in hydro code simulations of impact and implications for transient crater growth in porous targets. *Icarus* **180**, 514–527 (2006).
- Kemp, A. I. S. *et al.* Hadean crustal evolution revisited: New constraints from Pb–Hf isotope systematics of the Jack Hills zircons. *Earth Planet. Sci. Lett.* **296**, 45–56 (2010).
- Tonks, W. B. & Melosh, H. J. Magma ocean formation due to giant impacts. *J. Geophys. Res.* **98**, 5319–5333 (1993).
- Reese, C. C. & Solomatov, V. S. Fluid dynamics of local martian magma oceans. *Icarus* **184**, 102–120 (2006).
- Elkins-Tanton, L. T. & Hager, B. H. Giant meteoroid impacts can cause volcanism. *Earth Planet. Sci. Lett.* **239**, 219–232 (2005).
- Watters, W. A., Zuber, M. T. & Hager, B. H. Thermal perturbations caused by large impacts and consequences for mantle convection. *J. Geophys. Res. Planets* **114**, E02001 (2009).
- Potter, R. W. K., Collins, G. S., Kiefer, W. S., McGovern, P. J. & Kring, D. A. Constraining the size of the South Pole–Aitken basin impact. *Icarus* **220**, 730–743 (2012).
- Kamber, B. S., Whitehouse, M. J., Bolhar, R. & Moorbath, S. Volcanic resurfacing and the early terrestrial crust: Zircon U–Pb and REE constraints from the Isua

- Greenstone Belt, southern West Greenland. *Earth Planet. Sci. Lett.* **240**, 276–290 (2005).
23. Mojzsis, S. J., Harrison, T. M. & Pidgeon, R. T. Oxygen-isotope evidence from ancient zircons for liquid water at the Earth's surface 4,300 Myr ago. *Nature* **409**, 178–181 (2001).
 24. Wilde, S. A., Valley, J. W., Peck, W. H. & Graham, C. M. Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. *Nature* **409**, 175–178 (2001).
 25. Wichman, R. W. & Schultz, P. H. Sequence and mechanisms of deformation around the Hellas and Isidis impact basins on Mars. *J. Geophys. Res.* **94**, 17333–17357 (1989).
 26. Bedard, J. H. A catalytic delamination-driven model for coupled genesis of Archaean crust and sub-continental lithospheric mantle. *Geochim. Cosmochim. Acta* **70**, 1188–1214 (2006).
 27. Cavosie, A. J., Valley, J. W. & Wilde, S. A. Magmatic $\delta^{18}\text{O}$ in 4400–3900 Ma detrital zircons: a record of the alteration and recycling of crust in the Early Archaean. *Earth Planet. Sci. Lett.* **235**, 663–681 (2005).
 28. Abramov, O., Kring, D. A. & Mojzsis, S. J. The impact environment of the Hadean Earth. *Chemie Erde Geochem.* **73**, 227–248 (2013).
 29. Griffin, W. L. *et al.* The world turns over: Hadean–Archaean crust–mantle evolution. *Lithos* **189**, 2–15 (2014).
 30. Watson, E. B. & Harrison, T. M. Zircon thermometer reveals minimum melting conditions on earliest Earth. *Science* **308**, 841–844 (2005).
 31. Darling, J., Storey, C. & Hawkesworth, C. Impact melt sheet zircons and their implications for the Hadean crust. *Geochim. Cosmochim. Acta* **73** (Suppl.), 927–930 (2009).
 32. Holden, P., Lanc, P. & Ireland, T. *et al.* Mass-spectrometric mining of Hadean zircons by automated SHRIMP multi-collector and single-collector U/Pb zircon age dating: the first 100,000 grains. *Int. J. Mass Spectrom.* **286**, 53–63 (2009).
- Acknowledgements** We thank H. J. Melosh, S. J. Mojzsis, T. M. Harrison, A. J. Cavosie, A. I. S. Kemp, W. L. Griffin, M. M. Wielicki, R. J. Walker, E. B. Watson, P. Holden, O. Abramov, R. M. Canup, H. F. Levison, D. Nesvorný, O. Nebel, N. H. Sleep and N. Arndt for comments and criticisms that helped to shape the current paper. We gratefully acknowledge the developers of iSALE-2D/3D (www.isale-code.de). S.M., W.F.B. and D.A.K. received support from the NASA Solar System Exploration Research Virtual Institute grant no. NNA14AB03A and NNA14AB07A; K.W. and M.B. were supported by the Helmholtz-Gemeinschaft Deutscher Forschungszentren e.V. Alliance 'Planetary Evolution and Life'. A.M. was supported by the European Research Council Advanced Grant 'ACCRETE' (contract number 290568).
- Author Contributions** S.M. conceived the paper, built the Monte Carlo code and executed the simulations. W.F.B. and A.M. contributed to calibration and testing of the code, and to performing the fit of zircon age distributions. M.B. and K.W. executed iSALE impact simulations and processed the results. L.T.E.T. performed the geophysical interpretation of the output of the simulations. D.A.K. interpreted the melt volume. All authors contributed to the discussion of the results and their implications and to the crafting of the manuscript.
- Author Information** Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.M. (marchi@boulder.swri.edu).

METHODS

Impactor size–frequency distribution and flux. The SFD of bodies colliding with the early Earth cannot be directly constrained because subsequent geological evolution has erased the signatures of those impacts. Instead, we turn to very old cratered terrains found on the Moon, Mars and Mercury. The crater SFDs observed on those terrains has been used to constrain the shape of the impactor SFD in ancient times. The earliest visible populations of craters have a characteristic SFD resembling that of the current main asteroid belt^{12,33–36}. In this work we assume that a similarly shaped impactor SFD was striking the early Earth. Although it is natural to assume that the Earth and Moon have been exposed to the same impactor flux, a critical aspect of work concerns the impactor SFD at large sizes. In fact, the larger geometrical cross-section of the Earth than that of the Moon may allow larger projectiles to hit the Earth than the Moon. For comparison, the largest confirmed impact structure in the inner Solar System—the ~2,500-km South Pole–Aitken basin on the Moon—was produced by a projectile ~170 km across²¹. It is possible that larger objects struck the early Earth, in particular if the shape of the impactor SFD was shallow for large objects (see, for example, ref. 10). Therefore the impactor SFDs derived from crater populations observed on the terrestrial planets needs to be extrapolated to larger sizes if they are to be applied to the Earth. For the reasons discussed above, we considered a main-belt-like SFD up to ~1,000 km (Ceres), and we also considered an extension of the main-belt SFD up to 4,000 km (see the text and Extended Data Fig. 1 for further details).

The results of our work are not sensitive to the fine details of the shape of the impactor SFD. The more important issue is that the population of left-over planetesimals had large enough impactors (diameter >1,000 km) to allow it to reproduce the abundance of terrestrial HSEs. Here we consider a cut-off of 4,000-km and 1,000-km projectiles for impactors striking before and after 4.15 Gyr ago, respectively. The number of large objects in the population can conceivably be constrained by the largest lunar basin, provided that these impacts took place after the formation of the Moon's crust. Our modelling work indicates that the Earth was hit by 25–45 impactors larger than 200 km. Assuming a Earth-to-lunar scaling of ~20:1 (ref. 13), we obtain one or two South Pole–Aitken-forming impactors hitting the Moon. This is consistent with the lunar basin record and the lunar HSEs. We also estimate that an average of five impactors larger than 500 km hit the Earth. This translates into a ~70% ($= 1 - 6/20$) probability that the Moon escapes these impacts. These numbers provide a sanity check that our assumed impactor SFD is compatible with available constraints.

Note also that in an alternative scenario for the origin of the HSEs found within the mantles of the Earth and Moon³⁷, it has been argued that a thin, dynamically cold disk of small bodies spread across the terrestrial planet region would allow the Earth to accrete much more mass than the Moon. This scenario requires that a very thin disk be maintained during the planet formation era until the giant impact that made the Moon take place—probably many tens of millions of years to perhaps 100 Myr after Ca, Al-rich inclusions formation. It is unclear to us how this disk avoided dynamical excitation from planetary perturbations for this long interval (well after the solar nebula had dissipated). Beyond this, concerning HSEs on or in the Moon, this scenario would produce high enrichment in HSEs in the lunar crust (which is not observed) and it does not explain how the HSEs within the small bodies would breach the crust to reach the Moon's mantle. There is also no explanation provided of how this scenario would produce the crater SFDs found on ancient lunar terrains³⁵.

Concerning the terrestrial impact rate, we considered several scenarios. The nominal model assumes the so-called lunar sawtooth bombardment profile for the Moon⁴ extrapolated to the Earth. For this, we took the lunar impactor flux (as derived from the observed number of craters as a function of time), defocused it for the lunar gravitational field and Earth's gravitational field at the current lunar orbit, and then applied the Earth's gravitational focusing. More specifically, the crater size–frequency distribution on ancient lunar terrains has been converted into projectiles assuming first, a crater-to-projectile size scaling law for hard-rock³⁸; second, a lunar impact velocity of 11 km s⁻¹ for the period between 4.15 and 4.5 Gyr ago; and third, a lunar impact velocity of 22 km s⁻¹ for the period between 3.5 and 4.15 Gyr ago. The factor of ~2 increase in the impact velocity comes from observations of lunar crater populations³⁵ and is in agreement with dynamical estimates of terrestrial planet accretion¹¹ and how projectiles in the inner Solar System may have reacted to late migration of the giant planets¹³. The corresponding terrestrial impact velocities are ~16 and ~25 km s⁻¹. The projectile flux and impact velocity were then rescaled to the Earth, assuming that impactors producing the oldest visible lunar craters were not affected by Earth's gravitational focusing. This approximation is valid for most of the Moon's orbital evolution except for the first few million years (ref. 39), which are not relevant to the work presented here. This procedure gives a flux scaling factor for the Earth and Moon of ~1.24 and ~1.90 (per unit surface), respectively, for the LHB and pre-LHB times for the assumed impact speed at infinity. As detailed in the text, our extrapolation is

done by rescaling the lunar flux to the Earth. Here we assume that the LHB occurred at ~4.15 Gyr ago (the median value of the 4.1–4.2-Gyr interval of acceptable values, as concluded in refs 4, 13). The corresponding lunar flux curve is shown by the red line in Extended Data Fig. 2. We also considered the case of a narrow intense spike of LHB impacts at 3.9 Gyr ago, roughly corresponding to the scenario discussed in ref. 40 (cyan curve). Recent work has suggested that this LHB is unlikely to fit constraints^{4,13,14,41}. However, this assumption is still adopted by some researchers. Both of our impact flux curves were obtained by requiring that the integral of the accreted mass on the Moon match the abundance of lunar HSEs inferred to exist in the Moon's mantle, appropriately corrected to take into account partial accretion⁴. Finally, for the sake of completeness, we also considered two scenarios that exclude the LHB. The first is a simple extrapolation of the nominal case up to 4.5 Gyr ago (black curve). Note that here the total mass accreted by the Moon exceeds that predicted by the HSEs by a factor 3–4; therefore ref. 4 concluded that this scenario is unlikely. To compensate for this, we also considered a rescaled flux, reduced by a factor of one-third, to match the HSE constraint (green curve). All flux curves start at 4.5 Gyr ago, which is the assumed time for the formation of the Moon; however, our results are insensitive to the exact timing of Moon formation. We discuss how these curves compare with terrestrial zircon data in the following.

Terrestrial budget of HSEs. Terrestrial impactors may not be fully accreted with implication for the delivery of HSEs. For instance, this may be the case for large grazing projectiles⁴². In addition, projectiles larger than ~2,000 km may sequester mantle HSEs into Earth's core⁴³, whereas high-resolution smoothed-particle hydrodynamics simulations show that as much as ~50% of the core of a large differentiated impactor may plunge into the Earth's core (R. M. Canup, personal communication, August 2013). Thus, in scenarios involving collisions with projectiles larger than ~2,000 km the accreted mass would be higher than that estimated by HSEs (see Fig. 1).

Moreover, is it possible that some terrestrial HSEs predate the formation of the Moon? In canonical giant impact events that make the Moon⁷, most of the Earth's mantle is molten or partly molten, thus facilitating the segregation of HSEs into the core. Recent models have pushed the giant collisions to higher energies with respect to the canonical models; therefore an efficient segregation of pre-giant impact HSEs is even more likely (R. M. Canup, personal communication, April 2014). Ref. 44 found that $\epsilon^{182}\text{W}$ (that is, the ratio of sample $^{182}\text{W}/^{184}\text{W}$ to the terrestrial standard value, in parts per 10⁴) of the Moon is significantly different with respect to that of the Earth. This can be explained if an amount of tungsten with a broadly chondritic isotope ratio was delivered in chondritic proportions with the HSEs. This would suggest that HSEs were mostly accreted after the formation of the Moon.

Simulations of large terrestrial impacts with iSALE. Investigating the thermal effects of the early bombardment history on Earth implies a detailed quantitative understanding of the consequences of hypervelocity impacts of cosmic bodies of given mass, composition, velocity, and angle of incidence. Hydrocode modelling may serve as the most accurate approach with which to estimate crater size and the amount of shock wave-induced heating and melting of crustal and mantle rocks as the result of a collision. However, given the large number of collisions produced on the early Earth, it is impossible to model each impact event individually. A parameterization of the relationship between the size of an impact event (projectile mass m , diameter d , impact velocity v and impact angle α) and the resulting crater diameter D , depth h , volume V , excavation depth d_{ex} , excavation volume V_{ex} and melt volume V_{melt} is required. Existing scaling relationships (for example, refs 45–54) are based on laboratory and numerical experiments, analytical considerations and observations of the Earth's and lunar crater records. Whether these scaling relationships can be extrapolated to the size of impactors several hundred of kilometres in diameter as those that occurred in the early history of the Earth is questionable and requires further analysis and/or modifications of existing scaling laws to confirm their applicability to the given problem.

We used the hydrocode iSALE (see ref. 15 and references therein) to conduct a series of two-dimensional numerical models of impacts with projectile diameters d ranging from 1 to 1,000 km and impact velocities v of 8.5–17 km s⁻¹ (a few runs for 4,000-km impactors were also performed). In all models the impactor is resolved by 50 cells per projectile radius, and we assume a dunitic composition with a density $\delta = 3,314 \text{ kg m}^{-3}$. We do not consider the impact angle α , which naturally can vary between 0° and 90°, with the most likely encounter at 45°, and modelled vertical impacts (90°) only on a cylindrical, axial-symmetric two-dimensional grid on a planar target surface. Both simplifications (vertical impacts and planar target surface) reduce the computational costs of an individual simulation significantly and thus permit detailed parameter studies based on a sufficient number of numerical models. To compensate for the lack of varying impact angles in our models we assume an impact velocity that corresponds to the vertical component of the velocity vector ($v_p = v \sin \alpha$), an often-used simplification to approximate oblique impacts by two-dimensional simulations^{48,55,56} that was originally suggested in ref. 57.

In the baseline scenario, we assume a layered target composed of a lithosphere 125 km thick consisting of a 30-km granitic crust and a 95-km dunitic upper mantle. Within the lithosphere, heat is transported by conduction, giving rise to a relatively steep temperature gradient from 20 °C at the surface to 1,427 °C at the transition from the upper to the lower mantle (the asthenosphere; Extended Data Fig. 3). The temperature gradient in the asthenosphere is adiabatic according to the assumption that heat is transported by convection. At a depth of 2,930 km we consider an iron core with a constant temperature of 2,727 °C. Because we assume pure iron and neglect alloy composition, the solidus is relatively high and the core is in a solid state; however, the rheological properties of the core do not affect our models, because the craters in the biggest impact events under consideration do not reach as deep.

The thermodynamic behaviour of matter in our models is simulated by the ANEOS (Analytic Equation of State; ref. 58) for granite⁵³ and dunite⁵⁹. ANEOS can only account for one phase transition; we therefore took into account the solid state transition expected to have the biggest effect on the total amount of melt production by shock heating. Because we do not consider latent heat of melting, our estimate of shock melting corresponds to an upper estimate (for further explanations see refs 53, 60, 61). The rheological model, the resistance of rocks against plastic deformation, is explained in detail in ref. 62. We do not account for temporary weakening of matter during crater formation by acoustic fluidization⁶³ as required to explain mid-size complex crater morphologies. All material properties and model parameters are listed in Extended Data Table 1. We also neglect the effects of target spherical symmetry, which are estimated to contribute less than ~20% to the volume of melt for the impactor sizes considered in our work¹⁷. Note that direct two-dimensional cylindrical iSALE simulations (equivalent to full three-dimensional simulations for head-on collisions) have shown that the target curvature is negligible for projectile-to-target size ratios up to ~0.2 (ref. 64), corresponding to a projectile of 2,500 km for the Earth, in agreement with analytical estimates¹⁷. This gives confidence in the validity of the analytical estimates. The analytically estimated error of ~20% for the largest projectiles is within the error of our model.

All models begin with the first contact between the projectile and the target (we neglect the presence of an atmosphere) and stop after the collapse of the transient crater. Our models include the structural uplift of matter but do not last until all material is settled and the final crater is reached. Primarily our simulations aim at the determination of crater size, excavation depth and the melt that is generated by the impact event.

Previous studies aimed at the computation of impact melt volumes usually considered shock-induced melting only. Melting of rocks during impact is the result of shock-wave compression and subsequent release. Shock compression is an irreversible process in which plastic work is done on the target material that remains in the rock as heat after subsequent isentropic release and can raise the temperature of the target above the melt temperature. To quantify the amount of impact-generated melt it is necessary to determine the volume of material that experiences a peak shock pressure in excess of the material's critical shock pressure for melting (P_c). The critical shock pressure for melting (or the corresponding entropy) for granite and dunite is a material property that can be measured by shock experiments (see, for example, ref. 65) and serves as an input parameter for ANEOS⁵³. Because the petrographical composition for a given rock type, such as dunite, may vary, the stated P_c values found in the literature range from ~91 to ~156 GPa depending on whether pure fosterite or peridotite composition and incipient or complete melting are considered, respectively. Extended Data Fig. 4a shows the impact melt production determined by hydrocode simulations for impacts on layered targets (granite and dunite; see above) in comparison with scaling relationships proposed by refs 17, 18, 51. Extended Data Fig. 4b shows the dependence of melt production on impact velocity. The melt volume is determined in our hydrocode simulations by Lagrangian tracer particles that experience shock pressures in excess of P_c . Each tracer represents the amount of matter in the computational cell where it was initially located in (see, for instance, ref. 61). Apparently, the melt volume varies at most by a factor of two depending on the chosen P_c (91 or 156 GPa, respectively), which we consider to be insignificant for the present study. Thus, a more accurate approach considering partial melting if the post-shock temperature is between solidus and liquidus as proposed in ref. 66 was not included in this study. The much lower P_c for granite (46 GPa for incipient melting and 56 GPa for complete melting⁵³) raises the total amount of melt for impactors <100 km in diameter more significantly; however, for very large impactors (>100 km in diameter) the total amount of melt is dominated by mantle material, and the contribution of crustal material is negligible.

The critical pressure method for the quantification of impact melt production is in good agreement with estimates of the observed melt volumes at terrestrial impact craters⁴⁹, but it may not provide accurate estimates for very large impactors several hundred kilometres in diameter penetrating deep into Earth's mantle, for

two reasons. First, with increasing depth, where material experiences shock compression the pre-impact temperature and lithostatic pressure become important. The initial temperature of the rock affects the critical melt pressure P_c : preheated rock tends to show shock-metamorphic effects including melting at lower shock pressures than rocks at normal surface temperature^{67,68}. Second, at a depth approximately larger than the transition from the lithosphere to the asthenosphere, rocks may not melt at all because the lithostatic pressure raises the solidus above the shock-induced temperature increase. Contrarily, structural uplift of originally deep-seated material as a result of the gravity-driven collapse of the transient crater may give rise to decompression melting. However, the effect of decompression melting has been estimated to be small in comparison with shock-induced melting for impactors 20 km in diameter⁶⁹, but may well be important for very large impactors and steep geotherms⁶⁶.

To account for both effects (temperature increase with depth and unloading from the shock pressure to the lithostatic pressure at given depth) we used an alternative approach to determine the total amount of impact-generated melt. We simply record through all computational time steps in our hydrocode simulations whether the temperature of a tracer is in excess of the solidus temperature as a function of pressure for the given location (depth) and mark it as molten. For small impactors (a few kilometres in diameter) this approach provides the same results as the P_c method (see Extended Data Fig. 4). With increasing projectile diameter the melt volume deviates from the scaling lines (open circles) according to the critical pressure method, and slightly increased melt volumes occur (50–100 km projectile diameter; note the small variation at 25 km diameter resulting from the change from a granitic to a dunitic melt composition). For larger impactors (>100 km diameter) the melt volumes decrease below the expected trend according to a straight line on a double-logarithmic plot (power-law scaling). This is due to the fact that shock-heated material does not unload to pressures at which the post-shock temperature is in excess of the solidus temperature. The reason for this is the fact that the increase in solidus temperature with depth is steeper than the increase in adiabatic temperature in the lower mantle (Extended Data Fig. 3). Increasingly higher shock pressures are therefore required to raise the temperature above the solidus. In summary, the difference between iSALE simulations and the analytical estimates can be understood by accounting for the different assumptions used for each case. For example, using projectile diameters in the range ~100–1,000 km, our simulations predict more melt than analytical estimates do^{17,18}, because the code accounts for decompression melting. In contrast, for projectiles larger than 1,000 km, iSALE finds less melt than our analytical estimates^{17,18} because the latter neglects the increase of the solidus temperature as a function of depth.

Our analytical scaling relationships for melt production and crater sizes have been used as a rough guide, but they do not affect the conclusions of our paper; they are based on hydrocode modelling of impacts. Similarly, details of the composition and nature of the Hadean crust and the geotherm, for example, have little effect on our results. This is mainly because our conclusions are based on the effects of large collisions (impactors larger than 100 km); the volumes of melt produced by big impactors are largely insensitive to variation in the mantle potential temperature and its composition. The key factor is the overburden pressure, as detailed above.

We also find that the melt volume increases by ~75% if the mantle potential temperature increases by 200 °C (ref. 70) with respect to the standard case (or 1,600 °C). We also tested the case of a thinner lithosphere (80 km) and found that, for impactors larger than 1,000 km, the volume of melt is generally a few per cent higher than the nominal lithosphere (125 km), and up to 50% higher for impactor sizes between 100 and 1,000 km.

Finally, our estimates of impact-generated melt volume do not consider adiabatic melting of rising mantle elements. The latter process, yet to be the subject of a systematic study, is potentially very important for large projectiles. As discussed in ref. 20, projectiles larger than ~800 km may produce long-lasting perturbations to mantle dynamics. On a short timescale (1–10 Myr relevant for our work), the perturbation is characterized by umbrella-shaped patterns with rising elements at the centre, lateral spreading in the upper mantle and subsequent downwelling (see, for instance, Fig. 7 of ref. 20). As a result, voluminous quantities of rising mantle may melt adiabatically. It is therefore possible that our estimates of the melt production (Fig. 3) are lower bounds, in particular for projectiles larger than 1,000 km.

Concerning the estimate of the final crater size discussed in the text (see, for example, Fig. 2a), we relied on a transient-to-final crater size scaling law derived from lunar craters⁷¹. Recent work has found that the temperature of the lithosphere has an important role in the modification stage of crater formation²¹, and for a hotter target the final crater size can be twice as large as in classical cold scaling. The hot lithosphere scaling is probably more realistic for the early Earth; therefore the final crater sizes discussed here are likely to be underestimated by a factor of ~2. **Impact-generated melt extrusion and Hadean zircon formation.** Although it is well-known that zircons can crystallize from impact melt pools (see, for example,

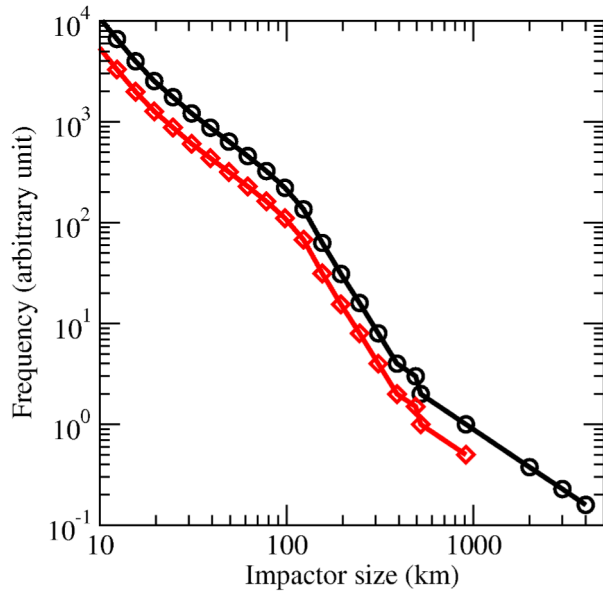
ref. 31), it is argued that the low crystallization temperatures of Hadean zircons are largely incompatible with such an origin^{30,31,72}. Our model shows that a fraction of the Earth's uppermost surface layer was melted by impacts. By analogy to recent terrestrial craters (for example Sudbury), it is expected that zircons should have also crystallized from the melt, even for intermediate-to-mafic melts via fractional crystallization⁷². This process would produce zircons that crystallized at higher temperatures than most Hadean zircons^{30,73}. Although it is conceivable that a fraction of the high-temperature Hadean zircons ($720^{\circ}\text{C} < T < 750^{\circ}\text{C}$) may come from impact melt pools (see, for example, ref. 28), this process seems secondary. Thus, a natural question arises: if impact melt was widespread on the Hadean Earth, why do Hadean zircons have a low crystallization temperature?

Our proposed mechanism naturally explains this observation. We computed the total volume of impact-generated melt produced in a spherical shell 100 km thick (that is, the crust) by all projectiles larger than 50 km. Larger projectiles were assumed not to contribute here because they blast through the shell. The result is that at least 10–20% of the shell is directly melted by impacts. For comparison, we also compute the volume of the shell buried by melt extrusion around the impacts (corresponding to an annulus from 10 to 30 times impactor radii from the impact point, only for projectiles larger than 100 km). This yields >800% of the shell volume, implying that the shell is reprocessed over and over. Therefore the latter process dominates by almost two orders of magnitude with respect to direct impact melt of the shell, explaining the paucity of high crystallization temperature among Hadean zircons. Note that this result is basically independent of the assumed thickness of the shell. A more significant contribution to zircon formation may come from the thick layers of melt extruded onto the surface (Fig. 2b and Extended Data Fig. 5). Given the deep origin of these magmas, however, they were probably ultramafic in composition, thus inhibiting significant zircon formation. Even assuming that low-temperature zircons could have formed from high-temperature melts as a result of fractional crystallization after a substantial decrease in temperature (see, for example, ref. 74), the volume of this fractionated reservoir would have been negligible with respect to the volume of the material buried. Thus the main contribution of these mantle mafic melts was probably as a crustal heat source, not as zircon source material. As discussed in the text, Hadean zircons probably crystallized from wet eutectic melts³. Such conditions may have been achieved as a result of impact-generated melt burial of large portions of the surface. In other words, buried weathered material could have been efficiently heated to wet melting conditions by magmatic intrusions (ref. 75, p. 168) from the mantle (as claimed to explain the volcanic plains around the 2,300-km Hellas basin on Mars; refs 19, 25, 76), and by the steepened geotherm produced by thinning of the lithosphere (close to the crater's rims). At larger radial distances, the sinking of lava and crustal recycling (see Fig. 10 of ref. 26; see also a recent commentary in ref. 77) may have been particularly important in a regime characterized by a highly fractured crust (due to impacts), as in the Hadean.

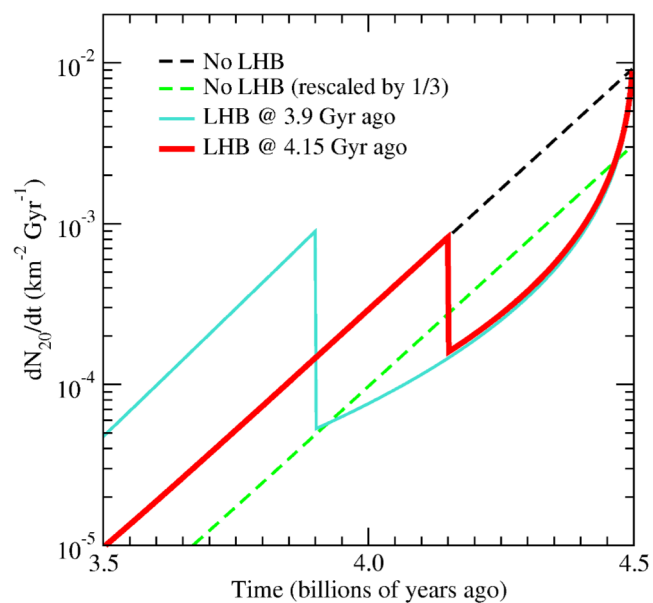
Finally, as discussed in the text, we investigated several impactor fluxes (Extended Data Fig. 2) and how they compare to the Hadean zircon age distribution. The key parameter for this comparison is the ratio of the diameter of the surface melt to the diameter of the projectile (f ; see Fig. 3). In the limits of the approximations described above (namely $f \approx 20$ –30; see the text), we find that only the nominal case (namely LHB at 4.15 Gyr ago; red curve in Extended Data Fig. 2) reproduces the Hadean zircon age distribution. The case of an LHB at 3.9 Gyr ago fails to reproduce the zircon data for any value of f , whereas the case with no LHB (green curve in Extended Data Fig. 2) only matches the zircon data for $f > 50$. This large f value is unjustified according to our estimates of melt volumes; this scenario is therefore extremely unlikely.

33. Strom, R. G., Malhotra, R., Ito, T., Yoshida, F. & Kring, D. A. The origin of planetary impactors in the inner Solar System. *Science* **309**, 1847–1850 (2005).
 34. Marchi, S., Mottola, S., Cremonese, G., Massironi, M. & Martellato, E. A new chronology for the Moon and Mercury. *Astron. J.* **137**, 4936–4948 (2009).
 35. Marchi, S., Bottke, W. F., Kring, D. A. & Morbidelli, A. The onset of the lunar cataclysm as recorded in its ancient crater populations. *Earth Planet. Sci. Lett.* **325**, 27–38 (2012).
 36. Fassett, C. I., Head, J. W. & Kadish, S. J. *et al.* Lunar impact basins: stratigraphy, sequence and ages from superposed impact crater populations measured from Lunar Orbiter Laser Altimeter (LOLA) data. *J. Geophys. Res. Planets* **117**, E00L08 (2012).
 37. Schlichting, H. E., Warren, P. H. & Yin, Q.-Z. The last stages of terrestrial planet formation: dynamical friction and the late veneer. *Astrophys. J.* **752**, 1–8 (2012).
 38. Melosh, H. J. *Impact Cratering: A Geologic Process* (Oxford Monographs on Geology and Geophysics no. 11, Clarendon Press, 1989).
 39. Murray, C. D. & Dermott, S. F. *Solar System Dynamics* (Cambridge Univ. Press, 1999).
 40. Ryder, G. Mass flux in the ancient Earth–Moon system and benign implications for the origin of life on Earth. *J. Geophys. Res. Planets* **107**, E45022 (2002).
 41. Norman, M. D. & Nemchin, A. A 4.2 billion year old impact basin on the Moon: U–Pb dating of zirconolite and apatite in lunar melt rock 67955. *Earth Planet. Sci. Lett.* **388**, 387–398 (2014).

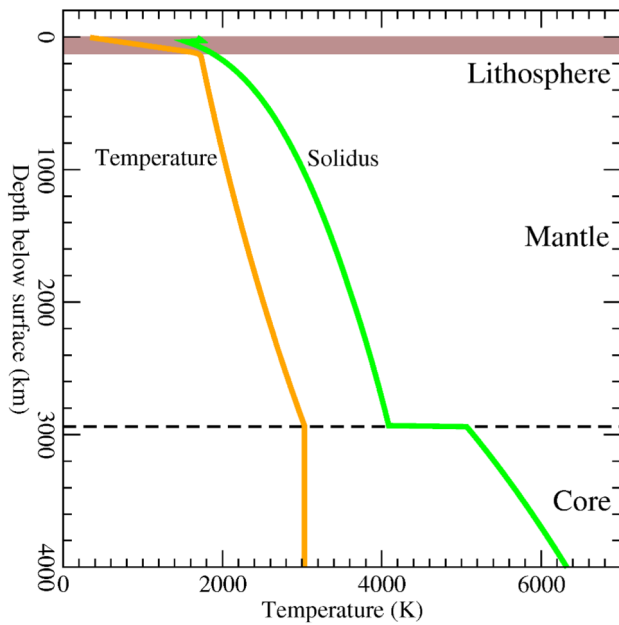
42. Leinhardt, Z. M. & Stewart, S. T. Collisions between gravity-dominated bodies. I. Outcome regimes and scaling laws. *Astrophys. J.* **745**, 79 (2012).
 43. Tonks, W. B. & Melosh, H. J. Core formation by giant impacts. *Icarus* **100**, 326–346 (1992).
 44. Kleine, T., Kruijer, T. S. & Sprung, P. in *Lunar and Planetary Science Conf.* 45 2895 (2014).
 45. O'Keefe, J. D. & Ahrens, T. J. Planetary cratering mechanics. *J. Geophys. Res.* **98** (E9), 17011–17028 (1993).
 46. Holsapple, K. A. The scaling of impact processes in planetary sciences. *Annu. Rev. Earth Planet. Sci.* **21**, 333–373 (1993).
 47. Wünnemann, K., Nowka, D., Collins, G. S., Elbeshausen, D. & Bierhaus, M. in *Proceedings of 11th Hypervelocity Impact Symposium*, 1–13 (2011).
 48. Elbeshausen, D., Wünnemann, K. & Collins, G. S. Scaling of oblique impacts in frictional targets: implications for crater size and formation mechanisms. *Icarus* 10.1016/j.icarus.2009.07.018 (2009).
 49. Grieve, R. A. F. & Cintala, M. J. An analysis of different impact melt-crater scaling and implications for the terrestrial impact record. *Meteoritics* **27**, 526–538 (1992).
 50. Ahrens, T. J. & O'Keefe, J. D. in *Impact and Explosion Cratering* (eds Roddy, D. J., Pepin, R. O. & Merrill, R. B.) 639–656 (Pergamon, 1977).
 51. Bjorkman, M. D. & Holsapple, K. A. Velocity scaling impact melt volume. *Int. J. Impact Eng.* **5**, 155–163 (1987).
 52. Grieve, R. A., Cintala, M. J. & Theriault, A. M. *Large-scale Impacts and the Evolution of the Earth's Crust: the Early Years* (Geol. Soc. Am. Spec. Pap. 405, 2006).
 53. Pierazzo, E., Vickery, A. M. & Melosh, H. J. A reevaluation of impact melt production. *Icarus* **127**, 408–423 (1997).
 54. Abramov, O., Wong, S. M. & Kring, D. A. Differential melt scaling for oblique impacts on terrestrial planets. *Icarus* **218**, 906–916 (2012).
 55. Ivanov, B. A. & Artemieva, N. A. in *Catastrophic Events and Mass Extinctions: Impact and Beyond* (eds Koeberl, C. & MacLeod, K.) 619–629 (Geol. Soc. Am. Spec. Pap. 356, 2002).
 56. Pierazzo, E. & Melosh, H. J. Melt production in oblique impacts. *Icarus* **145**, 252–261 (2000).
 57. Chapman, C. R. & McKinnon, W. B. in *Satellites* (eds Burns, J. A. & Matthews, M. S.) 492–580 (Univ. Arizona Press, 1986).
 58. Thompson, S. L. & Lauson, H. S. *Improvements in the Chart D Radiation—Hydrodynamic Code 3: Revised Analytic Equation of State* (Sandia Laboratories report SC-RR-71 0714, 1972).
 59. Benz, W., Cameron, A. G. W. & Melosh, H. J. The origin of the phase transition (g/cm^3) Moon and the single impact hypothesis. III. *Icarus* **81**, 113–131 (1989).
 60. Melosh, H. J. A hydrocode equation of state for SiO_2 . *Meteorit. Planet. Sci.* **42**, 2079–2098 (2007).
 61. Wünnemann, K., Collins, G. S. & Osinski, G. R. Numerical modelling of impact melt production in porous rocks. *Earth Planet. Sci. Lett.* **269**, 530–539 (2008).
 62. Collins, G. S., Melosh, H. J. & Ivanov, B. A. Modeling damage and deformation in impact simulations. *Meteorit. Planet. Sci.* **39**, 217–231 (2004).
 63. Melosh, H. J. Acoustic fluidization: a new geologic process? *J. Geophys. Res.* **84**, 7513–7520 (1979).
 64. Bierhaus, M., Noack, L., Wünnemann, K. & Breuer, D. in *Lunar and Planetary Science Conf.* 44, 2420 (2013).
 65. Stöffler, D. Deformation and transformation of rock-forming minerals by natural and experimental shock processes. I. Behavior of minerals under shock compression. *Fortschr. Mineral.* **49**, 50–113 (1972).
 66. Jones, A. P., Wünnemann, K. & Price, D. in *Plates, Plumes, and Paradigms* (eds Foulger, G. R., Natland, J. H., Presnall, D. C. & Anderson, D. L.) 711–720 (Geol. Soc. Am. Spec. Pap. 388, 2005).
 67. Huffman, A. R. & Reimold, W. U. Experimental constraints on shock-induced microstructures in naturally deformed silicates. *Tectonophysics* **256**, 165–217 (1996).
 68. Schmitt, R. T. Shock experiments with the H6 chondrite Kernouvé: pressure calibration of microscopic shock effects. *Meteorit. Planet. Sci.* **35**, 545–560 (2000).
 69. Ivanov, B. A. & Melosh, H. J. Impacts do not initiate volcanic eruptions: eruptions close to the crater. *Geology* **31**, 869–872 (2003).
 70. Herzberg, C., Condie, K. & Korenaga, J. Thermal history of the Earth and its petrological expression. *Earth Planet. Sci. Lett.* **292**, 79–88 (2010).
 71. McKinnon, W. B. & Schenk, P. M. Ejecta blanket scaling on the Moon and Mercury—inferences for projectile populations. *Lunar Planet. Inst. Sci. Conf. Abstr.* **16**, 544–545 (1985).
 72. Wielicki, M. M., Harrison, T. M. & Schmitt, A. K. Geochemical signatures and magmatic stability of terrestrial impact produced zircon. *Earth Planet. Sci. Lett.* **321**, 20–31 (2012).
 73. Harrison, T. M. & Schmitt, A. K. High sensitivity mapping of Ti distributions in Hadean zircons. *Earth Planet. Sci. Lett.* **261**, 9–19 (2007).
 74. Nutman, A. P. Comment on 'Zircon thermometer reveals minimum melting conditions on earliest Earth' II. *Science* **311**, 779 (2006).
 75. Turcotte, D. L. & Schubert, G. *Geodynamics* 2nd edn (Cambridge Univ. Press, 2002).
 76. Rogers, A. D. & Nazarian, A. H. Evidence for Noachian flood volcanism in Noachis Terra, Mars, and the possible role of Hellas impact basin tectonics. *J. Geophys. Res. Planets* **118**, 1094–1113 (2013).
 77. Pearce, J. A. Geochemical fingerprinting of the Earth's oldest rocks. *Geology* **42**, 175–176 (2014).
 78. Zahnle, K. *et al.* Emergence of a habitable planet. *Space Sci. Rev.* **129**, 35–78 (2007).
 79. Sleep, N. H. The Hadean–Archaean environment. *Cold Spring Harb. Perspect. Biol.* **2**, a002527 (2010).
 80. Arndt, N. T. & Nisbet, E. G. Processes on the young Earth and the habitats of early life. *Annu. Rev. Earth Planet. Sci.* **40**, 521–549 (2012).



Extended Data Figure 1 | Early Earth's impactor size–frequency distributions. The red curve corresponds to current main-belt asteroids larger than 10 km. The largest object is Ceres, whose diameter is ~ 913 km. The black curve (vertically shifted for clarity) is a replicate of the main-belt curve, extrapolated to 4,000 km by using the slope in the size range 500–913 km.

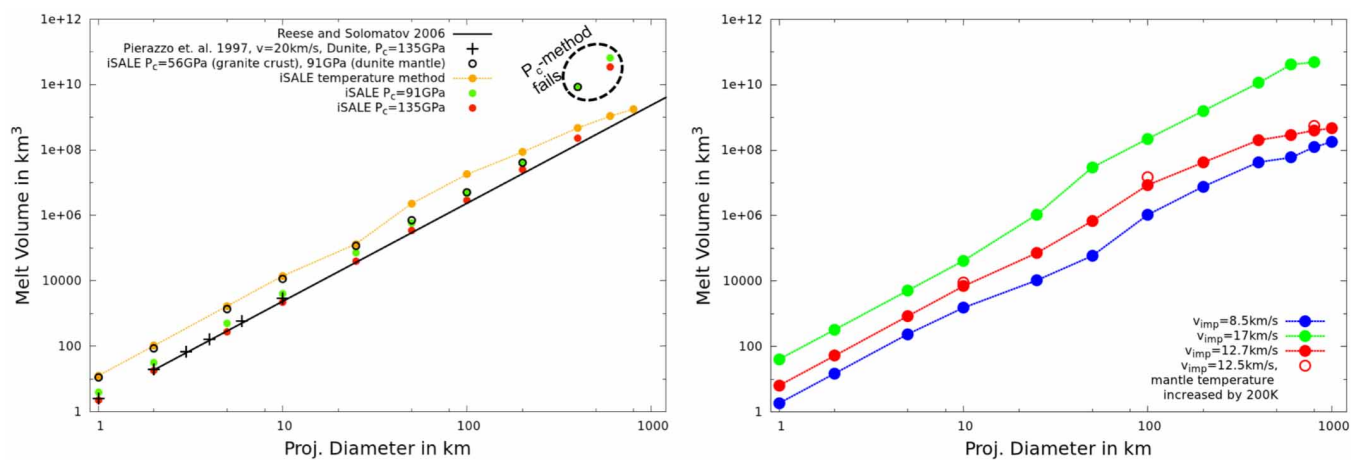


Extended Data Figure 2 | Lunar impact fluxes. The differential number of lunar craters >20 km (N_{20}) as a function of time and per unit surface for several scenarios discussed in the text.



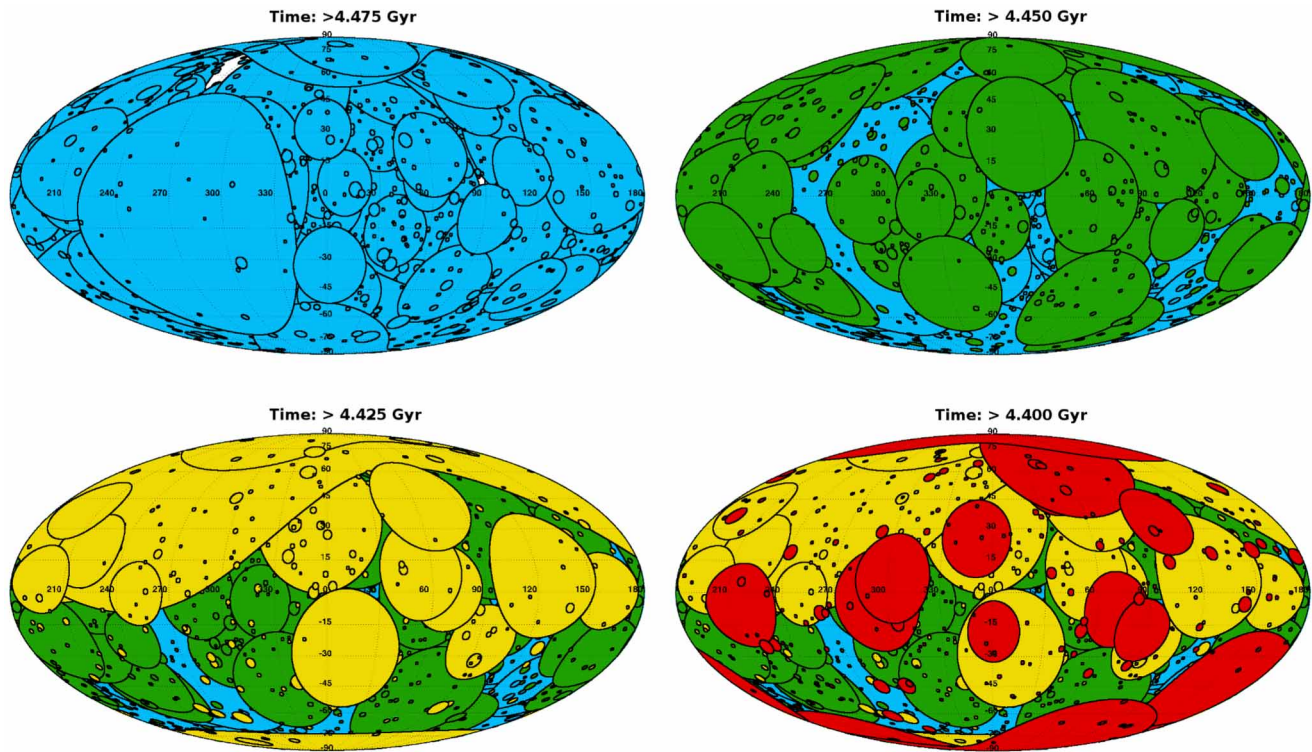
Extended Data Figure 3 | Solidus and geotherm used in iSALE simulations.

Note the temperature increase in the lithosphere that results in an increase in the temperature of buried surface material. Other processes resulting in an increase of the temperature of the buried crust are discussed in the text. The assumed thermal gradient is a lower limit (see Extended Data Table 1), implying that the increase in the temperature of the buried material can be significantly higher than is shown here.



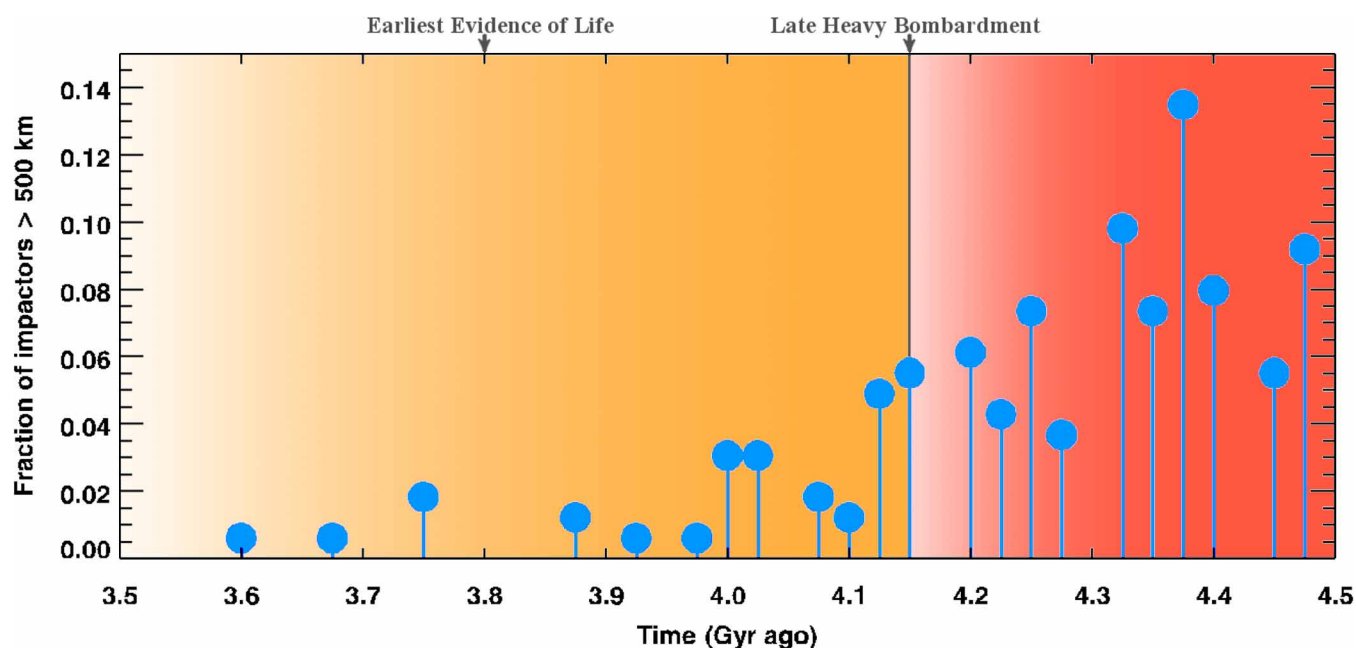
Extended Data Figure 4 | Impact-generated melt volume. Left, comparison of melt volume production for various methods; right, comparison of melt

volume production for various impact velocities and mantle potential temperature (see the text for more details).



Extended Data Figure 5 | Melt spreading over the first 100 Myr of Earth history. Mollweide projections of the cumulative record of craters at four different times. There are portions of the Earth's surface that are not affected by impact-generated melt at each time step, except for the first 25 Myr (or >4.475 Gyr). However, there is no significant fraction of the Earth's surface

that is unaffected by impacts before ~ 4 Gyr ago (see also Fig. 2b). Impacts therefore set the stage for the environmental conditions on the Hadean Earth and have implications for the origin and development of life (see, for instance, discussion in refs 78–80).



Extended Data Figure 6 | Minimum impact time for projectiles larger than 500 km. Blue dots indicate the minimum impact time for impactors larger than 500 km recorded in 163 successful Monte Carlo simulations (see Fig. 1). The vertical axis reports the number of impacts (in bins of 25 Myr each) normalized by the number of simulations. The lowest y values shown in the plot correspond

to one impact. The median time is 4.32 Gyr ago, and the mean is 4.27 Gyr ago. The earliest evidence of life on Earth (~ 3.8 Gyr ago), and the start of the Late Heavy bombardment (~ 4.15 Gyr ago; see the text) are also indicated. About 10% of the simulations have a minimum time of 4 Gyr ago or less.

Extended Data Table 1 | Various parameters used for iSALE simulations

Model parameters			
Cells per projectile radius (CPPR)	50		
Gravity (m/s ²)	9.81		
Crustal thickness (km)	30		
Mantle thickness (km)	2900		
Impact velocity (km/s)	8.5;12.7;17.0		
Impactor size (km)	1-4000		
Surface temperature (°C)	20		
Lithospheric temperature gradient (°C/km)	11.25		
Lithosphere thickness (km)	125		
Planet radius (km)	6371		
Core temperature (°C)	2727		

Material parameters			
	granite	dunite	iron
Melt temperature at zero pressure (°C)	1400	1100	1538
Heat capacity (J/(kg*K))	1000.0	1000.0	600.0
Constant Simon approximation for solidus (Pa)*	6.00E+9	1.52E+9	6.0E+9
Exponent in Simon approximation for solidus*	3.0	4.05	3.0
Cohesion of damaged material (Pa)	1.0E+4	1.0E+4	1.0E+4
Cohesion of intact material (Pa)	1.0E+7	1.0E+7	-
Coefficient of friction for intact material	0.6	0.6	0.4
Coefficient of friction for damaged material	2.0	1.2	-

Note that our results on impact-generated melt for large impactors are fairly insensitive to the assumed lithospheric geotherm gradient because most of the melt is produced at depth. Here we adopted a conservative low value. For additional discussion on the effects of geotherm see ref. 69.

* For the definition of these parameters, see ref. 69.

Seasonal not annual rainfall determines grassland biomass response to carbon dioxide

Mark J. Hovenden¹, Paul C. D. Newton² & Karen E. Wills¹

The rising atmospheric concentration of carbon dioxide (CO₂) should stimulate ecosystem productivity, but to what extent is highly uncertain, particularly when combined with changing temperature and precipitation¹. Ecosystem response to CO₂ is complicated by biogeochemical feedbacks² but must be understood if carbon storage and associated dampening of climate warming are to be predicted³. Feedbacks through the hydrological cycle are particularly important⁴ and the physiology is well known; elevated CO₂ reduces stomatal conductance and increases plant water use efficiency (the amount of water required to produce a unit of plant dry matter)⁵. The CO₂ response should consequently be strongest when water is limiting⁶; although this has been shown in some experiments⁷, it is absent from many^{8–11}. Here we show that large annual variation in the stimulation of above-ground biomass by elevated CO₂ in a mixed C₃/C₄ temperate grassland can be predicted accurately using seasonal rainfall totals; summer rainfall had a positive effect but autumn and spring rainfall had negative effects on the CO₂ response. Thus, the elevated CO₂ effect mainly depended upon the balance between summer and autumn/spring rainfall. This is partly because high rainfall during cool, moist seasons leads to nitrogen limitation, reducing or even preventing biomass stimulation by elevated CO₂. Importantly, the prediction held whether plots were warmed by 2 °C or left unwarmed, and was similar for C₃ plants and total biomass, allowing us to make a powerful generalization about ecosystem responses to elevated CO₂. This new insight is particularly valuable because climate projections predict large changes in the timing of rainfall, even where annual totals remain static¹². Our findings will help resolve apparent differences in the outcomes of CO₂ experiments and improve the formulation and interpretation of models that are insensitive to differences in the seasonal effects of rainfall on the CO₂ response^{7,13,14}.

Anthropogenic emissions of CO₂ have been increasing, reaching 9.7 billion tonnes of carbon (C) in 2012 (ref. 15). The biosphere absorbs approximately 50% of anthropogenically emitted carbon per annum¹⁶, so the continued ability of the biosphere to capture and sequester atmospheric CO₂ is a critical determinant of the atmospheric CO₂ concentration and thus of future climate. This is particularly true of the terrestrial biosphere where carbon sequestration seems to be increasing¹⁶. A principal uncertainty in projecting future sequestration is the extent of the elevated CO₂ (eCO₂) effect¹⁷ and, in particular, how this might be modified by potential changes in temperature and rainfall.

High CO₂ concentrations affect plant productivity in two main ways. First, elevated CO₂ (eCO₂) directly stimulates net carbon assimilation in C₃ plants by increasing carboxylation and suppressing photorespiration rates⁵. Second, eCO₂ lowers stomatal conductance thus reducing plant water use, leading to increased soil water retention⁵. These two physiological effects—increased photosynthesis and greater water use efficiency—lead to the widely held conclusion that the eCO₂ effect should be strongest when moisture is limited^{4,6,7,18,19}. If correct, this would be an extremely valuable generalization, which would provide much greater confidence in model projections of future CO₂ responses. Some experimental evidence does support the generalization as there are examples where the eCO₂ response was strongest in dry years^{4,7}; unfortunately

this not an invariable outcome and there are many studies that do not show a relationship between annual precipitation and the strength of the eCO₂ effect^{19,20}.

Demand for water in grasslands varies seasonally²¹, meaning that the benefit of an increased water use efficiency at eCO₂ would differ across seasons. In addition, the eCO₂ effect strengthens with increasing nitrogen availability^{6,19,22}, and nitrogen availability is strongly affected by the seasonal distribution of rainfall²³. It therefore seems possible, as suggested previously²⁴, that the eCO₂ effect might be related to seasonal rather than annual rainfall patterns. Here we test that possibility using data from a grassland exposed to elevated CO₂ and warming.

In the TasFACE Global Change Impacts Experiment²⁵, circular plots of a native grassland in southeastern Tasmania, Australia, were exposed to a factorial combination of eCO₂ concentration (control and 550 μmol mol⁻¹) and temperature (unwarmed and warmed by 2 °C). We assessed annual biomass production by harvesting in late summer (approximately February) over a 9-year period from 2002 to 2010. Because this site experiences summer drought, we define the year as starting 1 March, the first day of the southern autumn. This also matches this ecosystem's pattern of growth, which occurs mostly in spring and early summer. Rainfall during the experimental period was generally representative of the preceding century, although the period did contain the driest autumn and winters and the wettest spring on record (Extended Data Fig. 1). Because the strength of the eCO₂ effect varied from year to year, we tested several factors to gauge their influence on biomass stimulation by eCO₂. We used multimodel inference of multiple regression models to examine the importance of biomass in the control plots, biomass of the previous year, total annual rainfall, mean annual water availability, seasonal rainfall and mean seasonal water availability on the eCO₂ effect. Because the eCO₂ effect is strongly controlled by soil nitrogen¹⁹, we also tested the impact of soil mineral nitrogen content during spring.

Over the 9 years the eCO₂ effect averaged 13.9% but varied substantially among years (year × CO₂ $F_{8,64} = 3.67$, $P < 0.002$; Fig. 1), ranging from a suppression of biomass by $41.5 \pm 5.9\%$ (mean ± s.e.m.) to a stimulation of $96.3 \pm 25.0\%$ (mean ± s.e.m.; Fig. 1). The warming treatment did not vary across years (year × warming $F_{8,64} = 1.47$, $P = 0.18$) and importantly did not interact with the eCO₂ effect (CO₂ × warming $F_{1,8} = 0.01$, $P = 0.91$) nor did year, warming and eCO₂ interact (year × CO₂ × warming $F_{8,64} = 0.80$, $P = 0.61$). Therefore, warming by 2 °C did not alter the way that the eCO₂ effect varied across years.

The eCO₂ effect was poorly related to mean annual soil water availability ($r^2 = 0.27$, $F_{1,8} = 2.60$, $P = 0.15$) and annual rainfall ($r^2 = 0.42$, $F_{1,8} = 5.1$, $P > 0.05$) but could be predicted accurately from seasonal rainfall totals, with a model incorporating autumn, spring and summer rainfall totals describing 91% of the year-to-year variation in the eCO₂ effect ($r^2 = 0.91$, $F_{3,5} = 16.5$, $P < 0.005$). The predictive power of seasonal soil water availability (Fig. 1) was substantially lower than that of seasonal rainfall, suggesting that rainfall influences the eCO₂ effect in ways independent of its effect on soil water availability. Using weighted average values of the coefficients from all competitive models (Extended Data Table 1), the eCO₂ effect was found to be directly proportional to summer rainfall ($1.12 \pm 0.77\% \text{ mm}^{-1}$ (mean ± 95% confidence interval); Fig. 2) but

¹School of Biological Sciences, University of Tasmania, Hobart, Tasmania 7001, Australia. ²Land & Environmental Management, AgResearch, Palmerston North 4474, New Zealand.

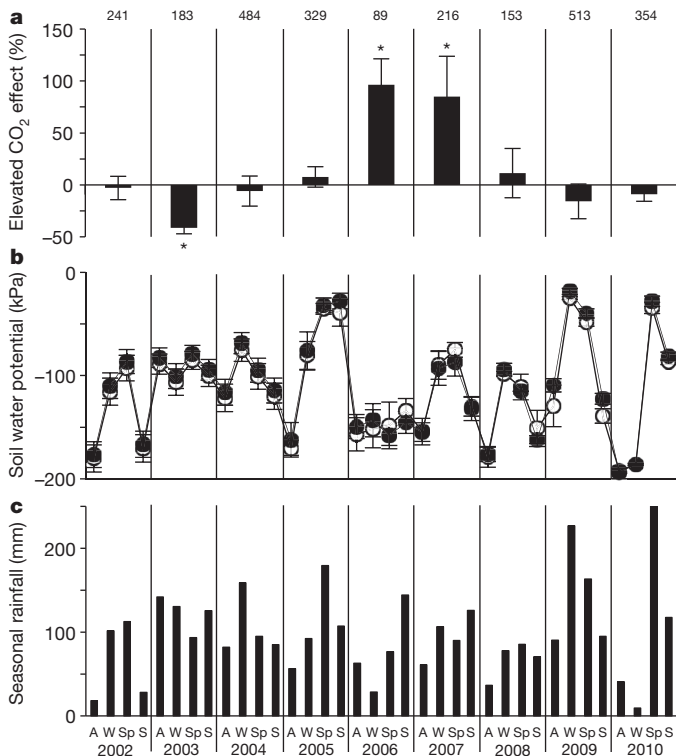


Figure 1 | Annual elevated CO₂ effect, seasonal soil water potential and rainfall in the TasFACE experiment. **a**, Average \pm s.e.m. ($n = 6$ replicate plots) percentage stimulation of total above-ground biomass by eCO₂ from 2002 to 2010. The eCO₂ effect was calculated as the percentage difference between elevated and ambient CO₂ plots compared with biomass in the ambient plots, pooled across both levels of warming treatment. A significant CO₂ effect for a particular year ($P < 0.05$) as determined by repeated measures analysis of variance is indicated by an asterisk. The mean above-ground biomass of the ambient plots is indicated by the numbers above each column. **b**, Average \pm s.e.m. ($n = 6$ replicate plots) seasonal soil water potential in elevated (filled symbols) and ambient CO₂ plots (empty symbols). **c**, Seasonal rainfall totals for each year starting 1 March. A, autumn; W, winter; Sp, spring; S, summer.

was reduced by increasing rainfall in both autumn ($-1.07 \pm 0.76 \text{ mm}^{-1}$ (mean \pm 95% confidence interval); Fig. 2) and spring ($-0.48 \pm 0.35 \text{ mm}^{-1}$ (mean \pm 95% confidence interval); Fig. 2). Variation in winter rainfall

had no influence on interannual variation in the eCO₂ effect as none of the competitive models included winter rainfall. Additionally, models that incorporated annual rainfall or soil water availability as a term were less accurate than simpler models including only seasonal rainfall. None of the other covariates examined improved the explanatory power of the models. Hence, the eCO₂ effect, whether combined with warming or not, depended upon the balance between summer and autumn/spring rainfall and was independent of other measured factors that varied among the years. The increase in productivity caused by eCO₂ was greatest in years when summer rainfall approached or exceeded the amount of spring and autumn rainfall (Fig. 3), and smallest or actually negative in years with abundant autumn/spring rainfall or little summer rainfall (Fig. 3).

The study system is a C₃–C₄ co-dominated grassland, so it is possible that the increased summer rainfall was only related to the eCO₂ effect on C₄ grasses, which are more active in the warmer summer months. This grassland contains a single C₄ grass species, which, although relatively abundant, was not always present in sufficient quantities to allow analysis. Therefore, we analysed the eCO₂ effect on C₃ biomass alone and found similar, but not identical, relationships with seasonal rainfall totals to that occurring with total biomass, with autumn and spring rainfall suppressing and summer stimulating the eCO₂ effect (Fig. 2). Although the seasonal rainfall balance did not predict the eCO₂ effect on C₃ biomass as well as it did for total biomass, the effect was similar and the rainfall balance still explained 75% of the variation in response across the years (Extended Data Fig. 2). Hence, the stimulatory effect of summer rainfall and the inhibitory effect of autumn/spring rainfall on the eCO₂ effect applied to both C₃-only and combined C₃ and C₄ vegetation, making these results widely relevant.

We believe these seasonal effects arose partly through differences in plant water use under eCO₂ but also through an interaction between rainfall amount and nitrogen availability. Our data show soil mineral nitrogen content declined in spring as rainfall increased (Fig. 4) and, because this was not accompanied by an increase in the amount of nitrogen located in plant biomass (Fig. 4), it is unlikely that the reduction was due to an increase in plant uptake but rather to leaching and/or denitrification; we have no data for autumn but assume similar processes would operate in this similar cool-wet season. As nitrogen availability declines it is well established that the eCO₂ effect also declines¹⁹, meaning that in spring/autumn there is a negative feedback on the eCO₂ response. As increasing rainfall in these seasons also reduces the effects of greater water use efficiency, the combined effects are to minimize eCO₂ responses in spring and autumn (Figs 2 and 3).

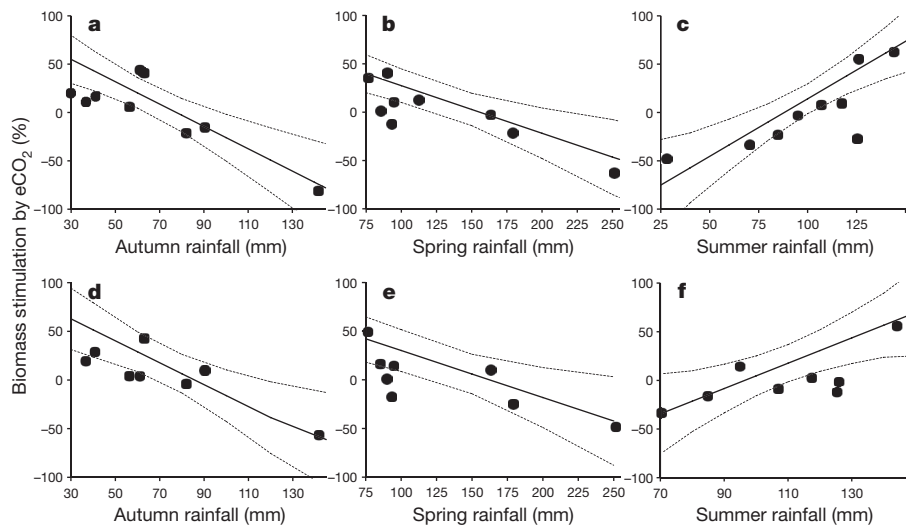


Figure 2 | The influence of seasonal rainfall variation on the elevated CO₂ effect. **a–f**, Partial regression plots showing the influence in the multiple regression model attributable to autumn (**a**, **d**), spring (**b**, **e**) and summer

(**c**, **f**) on the eCO₂ effect of total above-ground biomass (**a–c**) and C₃ above-ground biomass (**d–f**). The solid line shows the modelled effect, with 95% confidence limits shown as dashed lines.

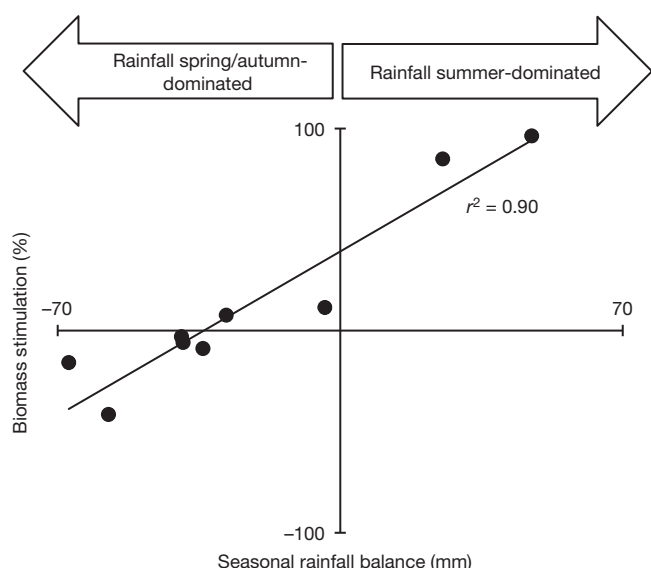


Figure 3 | The impact of seasonal rainfall balance on the elevated CO₂ effect. The mean ($n = 6$ replicate plots) annual eCO₂ effect on total above-ground biomass as a function of the seasonal rainfall balance, which is defined as the difference between summer rainfall and the sum of autumn and spring rainfall. Spring rainfall totals were halved in determining the seasonal rainfall balance as the multimodel estimates indicated that the effect of spring rainfall was approximately half that of the other seasons. Relatively more rainfall in summer gives a positive rainfall balance value, whereas a negative rainfall balance occurs when more rain falls in autumn and spring. The r^2 value was determined by linear regression ($F_{1,7} = 66.0$, $P < 0.0001$).

The situation differs in summer because rainfall in this and many other systems tends to occur in isolated, heavy events. In this case, growth stimulation by eCO₂ will occur because water savings lengthen the growing period from any single rainfall event. In between these summer downpours, eCO₂ will have little effect because the supply of water generally limits growth²⁰. An increase in the frequency of summer downpours increases the potential for increased growth and, therefore, the eCO₂ effect will increase with increasing summer rainfall. Because of these two opposing processes, stimulation of annual biomass by eCO₂ will depend upon the relative proportions of annual rainfall that occurs in

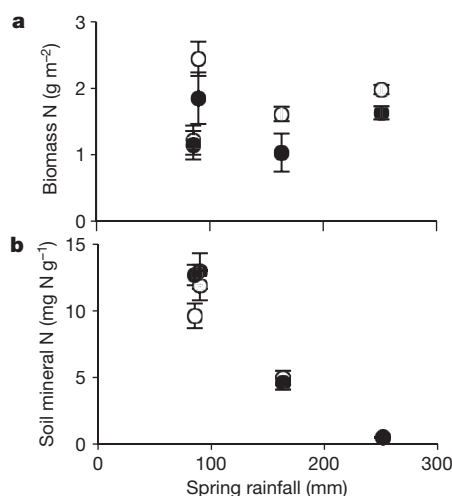


Figure 4 | The impact of spring rainfall on biomass nitrogen and soil mineral nitrogen content. **a**, Average \pm s.e.m. ($n = 6$ replicate plots) biomass nitrogen (N) and **b**, soil mineral nitrogen concentration in ambient (open symbols) and elevated CO₂ plots (filled symbols) during late spring as a function of total spring rainfall from 2007 to 2010. Soil mineral nitrogen concentration (sum of ammonium nitrogen and nitrate nitrogen concentrations) was obtained from 2 M KCl soil extracts.

autumn/spring and summer, and very little upon the total amount that falls over the year.

These results demonstrate that the pattern of rainfall during critical periods of the year has an overwhelming influence on the eCO₂ effect. Importantly, the actual impact of increasing rainfall on biomass stimulation by eCO₂ differs depending upon season; this might help explain the different patterns between annual rainfall and the eCO₂ effect in different experiments and among years in individual experiments. The results force us to reconsider the general view that the eCO₂ response of ecosystems will be strongest in drier conditions; this is a widespread view that has emerged from syntheses of experiments^{26,27}, has been the outcome of modelling^{14,28} and has consequently found its way into impact assessments^{29,30}. The interaction of water and nitrogen evident in our results has been identified as important previously^{6,29} but the flow-on consequences across seasons leading to diametrically opposite eCO₂ responses to rainfall is a finding that extends this understanding and suggests a review of models is necessary. Our data provide further experimental evidence that all global biogeochemical models used to model land–atmosphere interactions, particularly fluxes of carbon, need to include nutrients if they are to project fluxes accurately^{3,14}.

METHODS SUMMARY

The experiment was located in a species-rich native grassland dominated by C₃ grasses from the genus *Austrodanthonia* and the C₄ grass *Themeda triandra*. Full details of the experiment are available elsewhere²⁴. Every year above-ground biomass was sampled by clipping at the end of summer in one random 20 cm \times 20 cm quadrat in each of three randomly chosen quadrants of each circular plot. Above-ground material was separated to species, dried to constant weight at 60 °C and weighed. In October (mid-spring) 2007 to 2010, six soil cores (3.5 cm diameter, 10 cm depth) were collected from each plot. Soil mineral nitrogen content was assessed on 2 M KCl soil extracts as the sum of ammonium nitrogen and nitrate nitrogen.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 November 2013; accepted 25 March 2014.

Published online 28 May 2014.

- Intergovernmental Panel on Climate Change *Climate Change 2007 – The Physical Science Basis: Contribution of Working Group I to the Fourth Assessment Report of the IPCC* (eds Solomon, S. et al.) (Cambridge Univ. Press, 2007).
- Reich, P. B. & Hobbie, S. E. Decade-long soil nitrogen constraint on the CO₂ fertilization of plant biomass. *Nature Clim. Change* **3**, 278–282 (2013).
- Sokolov, A. P. et al. Consequences of considering carbon-nitrogen interactions on the feedbacks between climate and the terrestrial carbon cycle. *J. Clim.* **21**, 3776–3796 (2008).
- Morgan, J. A. et al. C₄ grasses prosper as carbon dioxide eliminates desiccation in warmed semi-arid grassland. *Nature* **476**, 202–205 (2011).
- Leakey, A. D. B. et al. Elevated CO₂ effects on plant carbon, nitrogen, and water relations: six important lessons from FACE. *J. Exp. Bot.* **60**, 2859–2876 (2009).
- McMurtrie, R. E. et al. Why is plant-growth response to elevated CO₂ amplified when water is limiting, but reduced when nitrogen is limiting? A growth-optimisation hypothesis. *Funct. Plant Biol.* **35**, 521–534 (2008).
- Morgan, J. A. et al. Water relations in grassland and desert ecosystems exposed to elevated atmospheric CO₂. *Oecologia* **140**, 11–25 (2004).
- Derner, J. D. et al. Above- and below-ground responses of C₃–C₄ species mixtures to elevated CO₂ and soil water availability. *Glob. Change Biol.* **9**, 452–460 (2003).
- Dukes, J. S. et al. Responses of grassland production to single and multiple global environmental changes. *PLoS Biol.* **3**, 1829–1837 (2005).
- Grunzweig, J. M. & Körner, C. Growth, water and nitrogen relations in grassland model ecosystems of the semi-arid Negev of Israel exposed to elevated CO₂. *Oecologia* **128**, 251–262 (2001).
- Marissink, M., Pettersson, R. & Sindhoj, E. Above-ground plant production under elevated carbon dioxide in a Swedish semi-natural grassland. *Agric. Ecosyst. Environ.* **93**, 107–120 (2002).
- Singh, D., Tsiang, M., Rajaratnam, B. & Diffenbaugh, N. S. Precipitation extremes over the continental United States in a transient, high-resolution, ensemble climate model experiment. *J. Geophys. Res.* **D 118**, 7063–7086 (2013).
- De Kauwe, M. G. et al. Forest water use and water use efficiency at elevated CO₂: a model-data intercomparison at two contrasting temperate forest FACE sites. *Glob. Change Biol.* **19**, 1759–1779 (2013).
- Piao, S. et al. Evaluation of terrestrial carbon cycle models for their response to climate variability and to CO₂ trends. *Glob. Change Biol.* **19**, 2117–2132 (2013).
- Peters, G. P. et al. The challenge to keep global warming below 2 °C. *Nature Clim. Change* **3**, 4–6 (2013).

16. Le Quere, C. *et al.* Trends in the sources and sinks of carbon dioxide. *Nature Geosci.* **2**, 831–836 (2009).
17. Matthews, H. D. Implications of CO₂ fertilization for future climate change in a coupled climate-carbon model. *Glob. Change Biol.* **13**, 1068–1078 (2007).
18. Körner, C., Morgan, J. & Norby, R. in *Terrestrial Ecosystems in a Changing World Global Change — The IGBP Series* (eds Canadell J. G., Pataki, D. E. & Pitelka, L. F.) Ch. 2, 9–21 (Springer, 2007).
19. Reich, P. B., Hungate, B. A. & Luo, Y. Q. Carbon-nitrogen interactions in terrestrial ecosystems in response to rising atmospheric carbon dioxide. *Annu. Rev. Ecol. Evol. Syst.* **37**, 611–636 (2006).
20. Norby, R. J. & Zak, D. R. Ecological lessons from Free-Air CO₂ Enrichment (FACE) experiments. *Annu. Rev. Ecol. Evol. Syst.* **42**, 181–203 (2011).
21. Knapp, A. K., Briggs, J. M. & Koelliker, J. K. Frequency and extent of water limitation to primary production in a mesic temperate grassland. *Ecosystems* **4**, 19–28 (2001).
22. Reich, P. B. *et al.* Nitrogen limitation constrains sustainability of ecosystem response to CO₂. *Nature* **440**, 922–925 (2006).
23. Borken, W. & Matzner, E. Reappraisal of drying and wetting effects on C and N mineralization and fluxes in soils. *Glob. Change Biol.* **15**, 808–824 (2009).
24. Leuzinger, S. & Körner, C. Rainfall distribution is the main driver of runoff under future CO₂-concentration in a temperate deciduous forest. *Glob. Change Biol.* **16**, 246–254 (2010).
25. Hovenden, M. J. *et al.* The TasFACE climate-change impacts experiment: design and performance of combined elevated CO₂ and temperature enhancement in a native Tasmanian grassland. *Aust. J. Bot.* **54**, 1–10 (2006).
26. Morgan, J. A. *et al.* CO₂ enhances productivity, alters species composition, and reduces digestibility of shortgrass steppe vegetation. *Ecol. Appl.* **14**, 208–219 (2004).
27. Nowak, R. S., Ellsworth, D. S. & Smith, S. D. Functional responses of plants to elevated atmospheric CO₂ – do photosynthetic and productivity data from FACE experiments support early predictions? *New Phytol.* **162**, 253–280 (2004).
28. Webb, N. P., Stokes, C. J. & Scanlan, J. C. Interacting effects of vegetation, soils and management on the sensitivity of Australian savanna rangelands to climate change. *Clim. Change* **112**, 925–943 (2012).
29. Polley, H. W. *et al.* Climate change and North American rangelands: trends, projections, and implications. *Rangeland Ecol. Manag.* **66**, 493–511 (2013).
30. Soussana, J. F. & Luescher, A. Temperate grasslands and global atmospheric change: a review. *Grass Forage Sci.* **62**, 127–134 (2007).

Acknowledgements This work was supported by grants from the Australian Research Council Discovery Projects scheme. We thank the Australian Department of Defence for access to the Pontville Small Arms Range Complex. J. Vander Schoor, A. Williams, J. Janes, J. Buettel and M. Porter provided technical assistance. R. Mallett assisted with statistical analyses and L. Barmuta provided advice on multimodel inference.

Author Contributions M.J.H. and P.C.D.N. conceived the TasFACE experiment, M.J.H. maintained and managed the experiment. K.E.W. assisted with the design and implementation of the biomass estimates and did preliminary analyses. M.J.H. and P.C.D.N. analysed the results and wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.J.H. (Mark.Hovenden@utas.edu.au).

METHODS

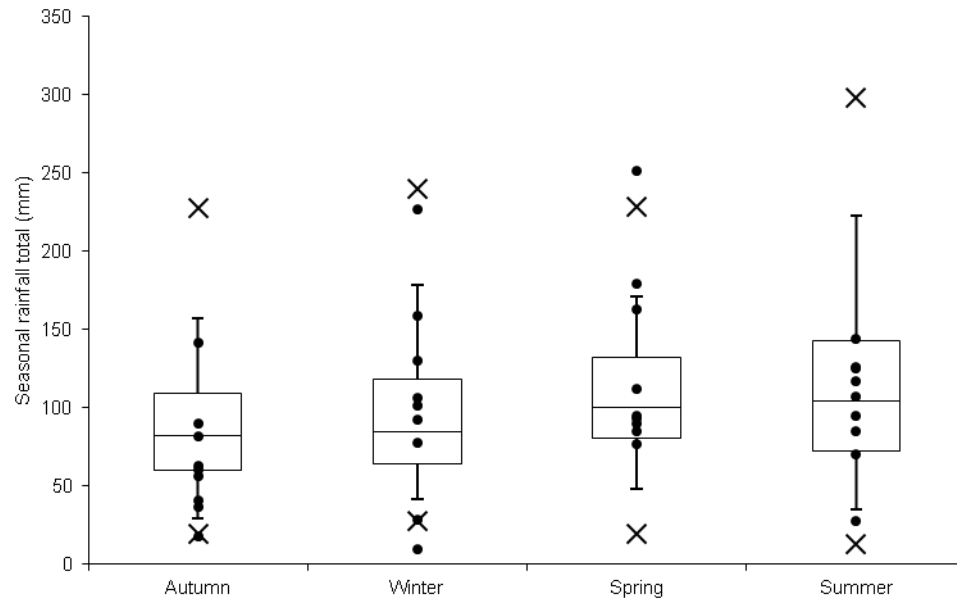
Experiment. The experiment was located in species-rich native temperate grassland in southeastern Tasmania, Australia (42°42' S 147°16' E). The region has a modified Mediterranean climate characterized by mild moist winters and warm dry summers. The vegetation was dominated by the perennial grasses *Austrodanthonia caespitosa*, *A. carphoides* and *T. triandra* (the only C₄ species) although almost one-third of the recorded species were native perennial forbs. The C₄ grass *T. triandra* contributes approximately 30% of the biomass in the grassland, although this varies substantially in space and from year to year. The experiment consisted of 12 free-air carbon dioxide enrichment (FACE) rings of 1.5 m diameter, in which vegetation was exposed either to ambient or elevated CO₂, and were either warmed or unwarmed. Thus, the experiment was a factorial 2 × 2 design with three replicate plots of each CO₂ × warming combination. FACE rings were fumigated to 550 μmol mol⁻¹ by the FACE method, using a modified pure-CO₂ injection system. Warming was provided by the addition of 140 W m⁻² of infrared radiation using 240 V 250 W Emerson Solid Ceramic Infrared Emitters suspended 1.2 m above the soil surface and above the centre of each ring. The infrared emitters operated continuously and provided an average warming of canopy temperature of 1.98 °C and of soil temperature at 1 cm depth of 0.82 °C over the year. Full experimental details are available elsewhere²⁴. Given the constraints of a manipulative experiment such as a FACE experiment, we calculated the probable effect size that we could detect given the number of experimental replicates and the background level of variation in biomass in this grassland. When a significant interaction between the two factors was not present, as was the case here, we calculated that we could detect a 30% change in site productivity.

Sampling and analyses. Every year, above-ground biomass was sampled by clipping to 2 mm above the ground surface at the end of summer (in February of the following calendar year) in one 20 cm × 20 cm quadrat randomly located in each of three randomly chosen quadrants of each circular plot. Above-ground material was separated to species, dried to constant weight at 60 °C and weighed. In October (mid-spring) 2007 to 2010, three soil cores (3.5 cm diameter, 10 cm depth) were collected from under three randomly chosen patches of each of the C₃- and C₄-dominated vegetation in each plot, giving six cores per plot. Soil cores were composited according to species in the field, returned to the laboratory upon ice and stored at 4 °C for a maximum of 2 days before being extracted. Soil ammonium nitrogen and nitrate nitrogen were determined for 2 M KCl extracts using protocols already described³¹.

Rainfall was measured using a Campbell Scientific tipping bucket rain gauge (0.2 mm event size) connected to a CR-10X data logger. Seasons were defined as 3-month periods starting on the first day of the month, with autumn starting on 1 March.

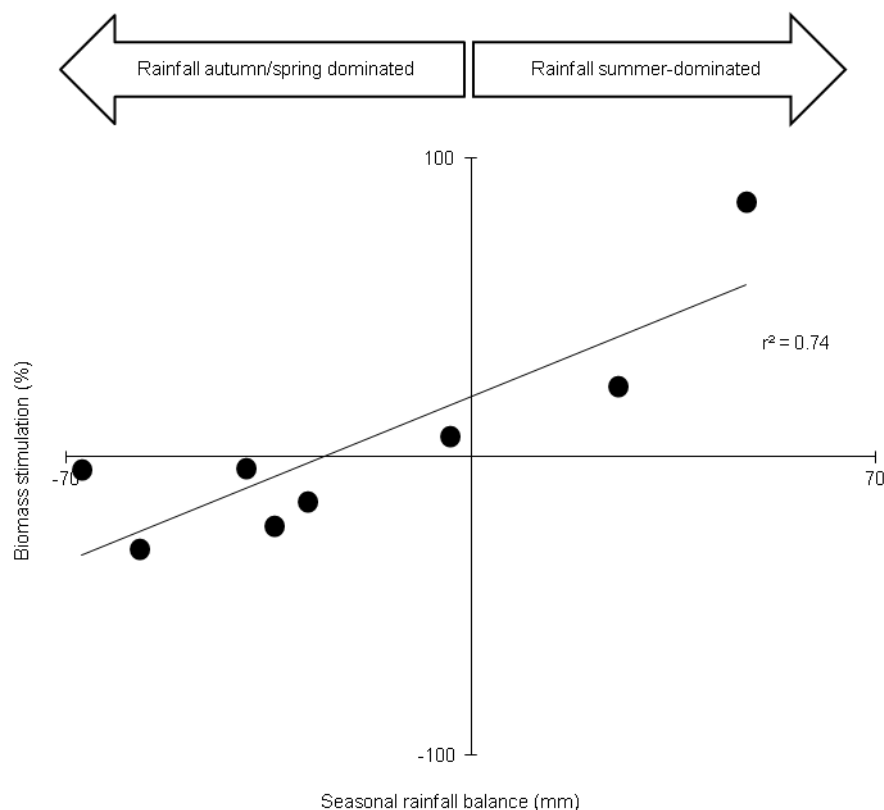
Statistical analyses. Biomass data were analysed using a repeated-measures two-factor analysis of variance (ANOVA) in R³², with CO₂ and warming as fixed factors, which was appropriate given the two-factor orthogonal design of the experiment. Biomass data were tested for normality and heteroscedasticity using box plots, residual plots and Cochran's test when these indicated heteroscedasticity was likely³³. Consequently, data were transformed to their natural logarithm before ANOVA, which provided a homoscedastic, normally distributed data set. As the analysis indicated that there were no CO₂ × warming or year × CO₂ × warming interactions, the CO₂ fertilization effect was determined on samples pooled across warming treatments. Thus, the percentage stimulation of biomass by eCO₂ was calculated as the difference in biomass between eCO₂ and ambient plots in proportion to the biomass of ambient plots, ignoring the warming treatment. Relationships between seasonal rainfall totals and the CO₂ fertilization effect were determined by multiple regression analyses using R. Potential collinearity between seasonal rainfall totals was examined by simple regressions and scatter plots³³. There was no evidence of any collinearity between seasonal rainfall totals, so all seasons were retained in the multiple regression analyses. Relationships between the CO₂ fertilization effect and seasonal rainfall totals were determined using multimodel inference procedures³⁴ with the MuMIn package in R. Beginning with all possible combinations of seasonal rainfall totals, we ranked the resultant models using the Akaike information criterion corrected for finite sample size (AIC_c). Model competitiveness was determined by observation of the difference in AIC_c between each model and the lowest value of AIC_c obtained (ΔAIC_c). Models were ranked in ascending ΔAIC_c value and a distinction between competitive and non-competitive models was made by observing any obvious breaks in the sequence of ascending ΔAIC_c. By this method models with ΔAIC_c < 6 were deemed competitive and those with ΔAIC_c > 10 non-competitive. Model term coefficients were determined by calculating weighted average values over all competitive models³⁴ and 95% confidence limits calculated. As the winter rainfall term was not included in any of the competitive models, it was omitted from further analysis. We used partial regression analysis to determine the effects, with 95% confidence limits, of autumn, spring and summer rainfall totals on the annual CO₂ fertilization effect for total above-ground biomass as well as for that of C₃ species alone using the effects package in R.

31. Osanai, Y. *et al.* Decomposition and nitrogen transformation rates in a temperate grassland vary among co-occurring plant species. *Plant Soil* **350**, 365–378 (2012).
32. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Development Core Team, 2011).
33. Quinn, G. P. & Keough, M. J. *Experimental Design and Data Analysis for Biologists* (Cambridge Univ. Press, 2002).
34. Burnham, K. P. & Anderson, D. R. *Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach* 2nd edn (Springer, 2002).



Extended Data Figure 1 | Seasonal rainfall during the study period compared with seasonal rainfall for the previous 100 years. The seasonal rainfall during the study period is shown by filled data points, with the mean and interquartile values shown in the box plot. Error bars indicate the fifth

and ninety-fifth percentile ranges; crosses indicate maximum and minimum recorded values from the previous 100 years. Historical values were obtained from the nearby Australian Bureau of Meteorology weather station at Bagdad.



Extended Data Figure 2 | The impact of seasonal rainfall balance on the elevated CO₂ effect on biomass of C₃ vegetation only. The mean ($n = 6$ replicate plots) annual elevated CO₂ effect on above-ground biomass of C₃ plants only as a function of the seasonal rainfall balance, which is defined as the difference between summer rainfall and the sum of autumn and spring rainfall. Spring rainfall totals were halved in determining the seasonal rainfall

balance as the multimodel estimates indicated that the effect of spring rainfall was approximately half that of the other seasons. Relatively more rainfall in summer gives a positive rainfall balance value, whereas a negative rainfall balance occurs when more rain falls in autumn and spring. The solid line and associated r^2 value are the result of a linear regression analysis ($F_{1,6} = 18.1$, $P < 0.006$).

Extended Data Table 1 | Model coefficients and performance for CO₂ effect on above-ground biomass

Model #	Autumn rainfall	Spring rainfall	Summer rainfall	df	AIC _c	ΔAIC _c	Weight
1	-1.16	-0.49	1.19	5	102.2	0	0.68
2	-0.90	-	1.00	4	104.4	2.11	0.23
3	-	-0.32	0.58	4	107.8	5.54	0.04
4	-0.54	-0.37	-	4	108.1	5.85	0.04

Models are ranked in order of the corrected Akaike information criterion (AIC_c), df, degrees of freedom. ΔAIC_c, the difference in AIC_c from the best model. Weight, Akaike weight, which corresponds to the probability that the model is the best of all models tested.

PTEX is an essential nexus for protein export in malaria parasites

Brendan Elsworth^{1,2*}, Kathryn Matthews^{3*}, Catherine Q. Nie¹, Ming Kalanon³, Sarah C. Charnaud^{1,2}, Paul R. Sanders¹, Scott A. Chisholm³, Natalie A. Counihan³, Philip J. Shaw⁴, Paco Pino⁵, Jo-Anne Chan¹, Mauro F. Azevedo¹, Stephen J. Rogerson⁶, James G. Beeson^{1,2,6}, Brendan S. Crabb^{1,2,6*}, Paul R. Gilson^{1,2*} & Tania F. de Koning-Ward^{3*}

During the blood stages of malaria, several hundred parasite-encoded proteins are exported beyond the double-membrane barrier that separates the parasite from the host cell cytosol^{1–6}. These proteins have a variety of roles that are essential to virulence or parasite growth⁷. There is keen interest in understanding how proteins are exported and whether common machineries are involved in trafficking the different classes of exported proteins^{8,9}. One potential trafficking machine is a protein complex known as the *Plasmodium* translocon of exported proteins (PTEX)¹⁰. Although PTEX has been linked to the export of one class of exported proteins^{10,11}, there has been no direct evidence for its role and scope in protein translocation. Here we show, through the generation of two parasite lines defective for essential PTEX components (HSP101 or PTEX150), and analysis of a line lacking the non-essential component TRX2 (ref. 12), greatly reduced trafficking of all classes of exported proteins beyond the double membrane barrier enveloping the parasite. This includes proteins containing the PEXEL motif (RxLxE/Q/D)^{1,2} and PEXEL-negative exported proteins (PNEPs)⁶. Moreover, the export of proteins destined for expression on the infected erythrocyte surface, including the major virulence factor PfEMP1 in *Plasmodium falciparum*, was significantly reduced in PTEX knockdown parasites. PTEX function was also essential for blood-stage growth, because even a modest knockdown of PTEX components had a strong effect on the parasite's capacity to complete the erythrocytic cycle both *in vitro* and *in vivo*. Hence, as the only known nexus for protein export in *Plasmodium* parasites, and an essential enzymic machine, PTEX is a prime drug target.

To address the role of PTEX in protein export directly, we examined parasite lines defective in PTEX components for their capacity to translocate exported proteins. Two PTEX components, TRX2 and PTEX88 (Fig. 1a), have auxiliary roles in PTEX function, because their deletion results in a substantial parasite growth defect^{12,13}. We therefore assumed that PTEX function is suboptimal in these lines. Here we show that surface expression of parasite antigens was substantially reduced in *Plasmodium berghei* TRX2-deficient parasites¹² (Extended Data Fig. 1), which is consistent with a role for PTEX in protein export. To perturb PTEX function more fully, conditional mutants of two essential PTEX components, HSP101 and PTEX150 (refs 10, 12, 13), were also generated. These proteins are synthesized in late schizogony and early ring stage and reside in the parasitophorous vacuole membrane for the remainder of the erythrocytic cycle¹⁴.

For HSP101, we generated a *P. berghei* line, Pbi101 KD, harbouring HSP101 under the transcriptional control of an anhydrotetracycline (ATc)-regulated transactivator element¹⁵ (Fig. 1b, c and Extended Data Fig. 2a). The growth of Pbi101 KD parasites was specifically sensitive to treatment with ATc (Fig. 1d, e). The Pbi101 KD line grew poorly in mice pre-exposed to ATc 24 h before infection (Fig. 1d, top panel, and Extended Data Fig. 2b), and normal growth of Pbi101 KD in the absence of ATc could be reversed

if ATc was added at day 4 (Fig. 1d, middle panel, and Extended Data Fig. 2b). As expected¹⁵, the growth of parental *P. berghei* ANKA parasites was unaffected by the presence of ATc (Fig. 1d, bottom panel).

To examine the growth effect in more detail, purified ring-stage Pbi101 KD parasites were injected into mice pre-exposed to ATc, then isolated 29 h later and cultured *in vitro* with ATc (Fig. 1e). As expected, parasites invaded erythrocytes in the mice and developed normally into ring stages (Fig. 1e, 24 h time point). However, parasites appeared morphologically abnormal by the 34 h time point and were incapable of developing into schizonts by the 46 h time point, unlike Pbi101 KD parasites not exposed to ATc. Asynchronous Pbi101 KD ring-stage parasites cultured *in vitro* for 16 h in the presence of ATc demonstrated a threefold to sixfold decrease in *hsp101* messenger RNA in schizont stages (Fig. 1f) and a 85–90% knockdown of HSP101 protein by the 29 h time point in Fig. 1e relative to the loading control proteins EXP2 and MSP8 (Fig. 1g).

To examine whether HSP101 knockdown affected protein export, we assessed whether asynchronous Pbi101 KD parasites harvested from ATc-pretreated mice at the time point indicated by the grey bar in Fig. 1d displayed surface-expressed antigens. Pbi101 KD parasites showed a strong reduction in parasite-encoded surface antigens compared with parasites grown in the absence of ATc (Fig. 2a). In an alternative approach, Pbi101 KD parasites grown to a higher parasitaemia (~10%) in the absence of ATc were subsequently treated with ATc for either 12 or 24 h before analysis. Given the asynchronicity of the infection, some parasites would already have transcribed HSP101 before ATc treatment commenced; consistent with this, surface expression was reduced in a manner dependent on the duration of ATc exposure (Fig. 2b).

We also assessed the export of individual proteins by immunofluorescence assay (IFA) using specific antibody reagents against three different exported proteins. Two of these proteins, PbANKA_114540 and PbANKA_122900, contain the PEXEL motif (RxLxE/Q/D) and localize to punctate structures in the erythrocyte cytosol¹⁶ (C. K. Moreira, B. Naissant, A. Coppi, L. Bennet, E. Aime, B. Franke-Fayard, C. J. Janse, I. Coppens, P. Sinnis and T. J. Templeton, personal communication), whereas PbANKA_083680 (EMAP1) is a PNEP that localizes to the erythrocyte membrane¹⁷. In each case, a striking blockage of protein export was observed in Pbi101 KD parasites exposed to a variety of different ATc treatment regimes; this included Pbi101 KD harvested from mice at the time points represented by the grey bar and asterisks in Fig. 1d (Fig. 2c and Extended Data Fig. 3a, respectively) and in Fig. 1e (Fig. 2c and Extended Data Fig. 3a). For example, in morphologically normal ring-stage parasites examined at the 24 and 29 h time points in Fig. 1e, observable export of PbANKA_114540 and PbANKA_083680 could only be detected in 7 out of 131 and 5 out of 100 parasites, respectively, whereas none of 100 parasites visibly exported PbANKA_122900. In contrast, almost equivalent numbers of Pbi101 KD parasites grown with or without ATc (80% and 88% expression, respectively) expressed MSP8,

¹Macfarlane Burnet Institute for Medical Research and Public Health, Melbourne, 3004, Australia. ²Monash University, Clayton, Victoria, 3800, Australia. ³Deakin University, Waurn Ponds, 3216, Australia. ⁴National Center for Genetic Engineering and Biotechnology (BIOTEC), Pathum Thani 12120, Thailand. ⁵The University of Geneva, 1211 Geneva 4, Switzerland. ⁶The University of Melbourne, Parkville, Victoria, 3010 Australia.

*These authors contributed equally to this work.

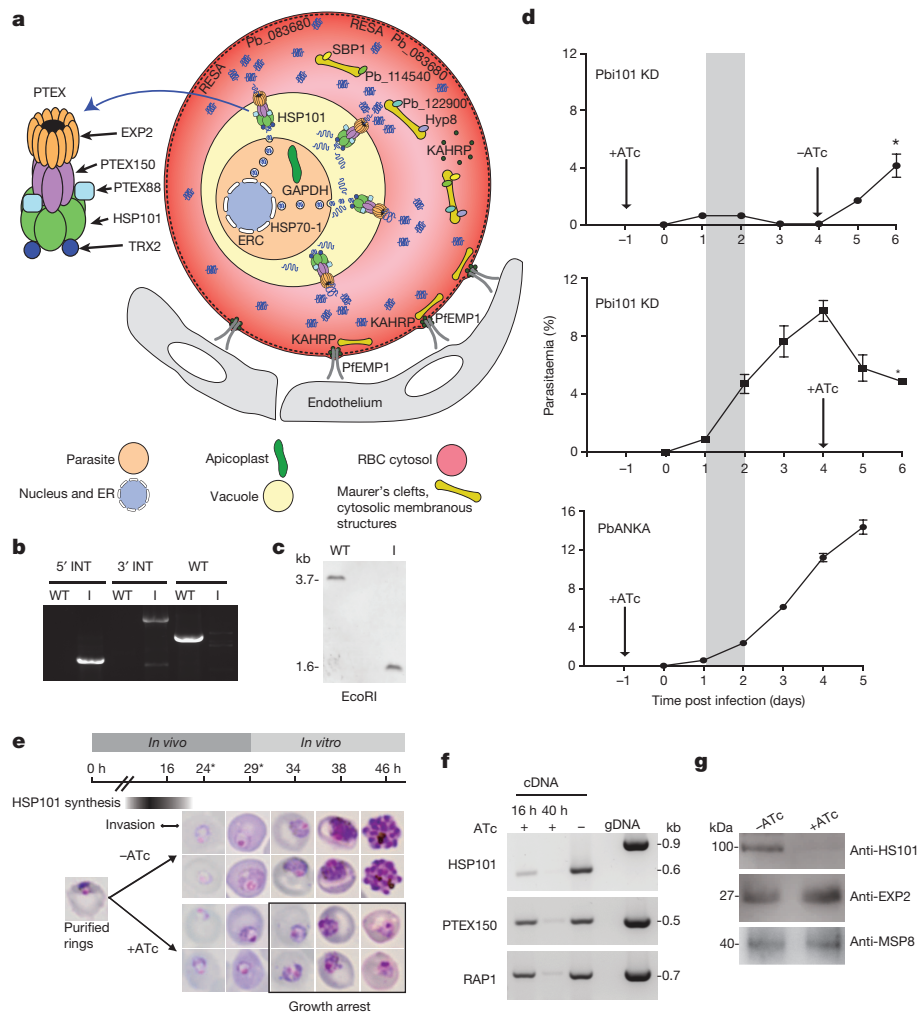


Figure 1 | Inducible knockdown of *P. berghei* HSP101 (i101 KD).

a, Diagram of a parasite-infected erythrocyte (RBC), the location of PTEX and proteins investigated in this study. ER, endoplasmic reticulum; GAPDH, glyceraldehyde-3-phosphate dehydrogenase. **b**, **c**, PCR (**b**) and Southern blot (**c**) of Pbi101 KD (I) and PbANKA wild-type (WT) parasites. kb, kilobases; INT, integration. **d**, Representative experiments ($n = 3$) showing that growth of Pbi101 KD *in vivo* is affected by ATc. Error bars show s.e.m. for three mice per

a parasite-membrane protein known to be strictly synthesized in the ring stage^{18,19} (Fig. 2d). The localization of additional control proteins, EXP2 and the apicoplast-resident protein (ACP), were also unaffected by HSP101 knockdown (Extended Data Fig. 3b). As expected, export was unaffected in wild-type parasites treated with ATc (Extended Data Fig. 3a). In summary, knockdown of HSP101 in *P. berghei* induced a profound defect in the capacity of both PEXEL-containing proteins and PNEPs to enter the cytosol of the infected erythrocyte, with a consequential detrimental effect on parasite growth, thereby highlighting an essential function for PTEX.

In a parallel approach, an inducible ribozyme system²⁰ was used to generate a conditional PTEX150 knockdown in the human malaria parasite *P. falciparum*. CS2 was used as the parental parasite strain because this line expresses a stable and well-characterized PfEMP1 phenotype^{7,21}. The gene encoding PTEX150 was modified to incorporate the glucosamine-inducible *glmS* ribozyme within its 3' untranslated region to generate a line termed PTEX150-HAGlmS (Extended Data Fig. 4). A control line, PTEX150-HA, identical to PTEX150-HAGlmS except for the absence of the *glmS* ribozyme, was also generated.

PTEX150-HAGlmS parasites exposed to glucosamine at the trophozoite stage (Fig. 3a, b, day 0) remained capable of invading erythrocytes normally (day 1). However, in contrast to the control line, they could

not advance beyond the early trophozoite stage and hence could not progress to the next parasite cycle when glucosamine concentrations of 0.6 mM or more were used (day 3) (Fig. 3b and Extended data Fig. 5). At glucosamine concentrations less than this, DNA replication and growth progressed but were slightly lower than in the control line, in a dose-dependent manner.

When glucosamine was added at the trophozoite stage, PTEX150 protein levels were reduced in PTEX150-HAGlmS ring-stage parasites in a glucosamine dose-dependent manner, with more than 50% knockdown at glucosamine concentrations above 0.3 mM (Fig. 3c and Extended Data Fig. 6a). In contrast, PTEX150 levels were unaffected in PTEX150-HA control parasites exposed to glucosamine. The smaller but reproducible *glmS*-specific decrease observed for the control protein endoplasmic reticulum-resident calcium-binding protein (ERC) could be explained by the strong PTEX150 knockdown-specific growth effect, which we expected would reduce the expression of ERC and other control proteins (Extended Data Fig. 6a, b). Together these data indicated that PTEX150 levels were specifically reduced by the addition of glucosamine to PTEX150-HAGlmS parasites, and that blood-stage development was exquisitely sensitive to PTEX150 levels, with 50% or more knockdown effectively ablating development to the mature trophozoite stage in the parasite cycle after glucosamine treatment.

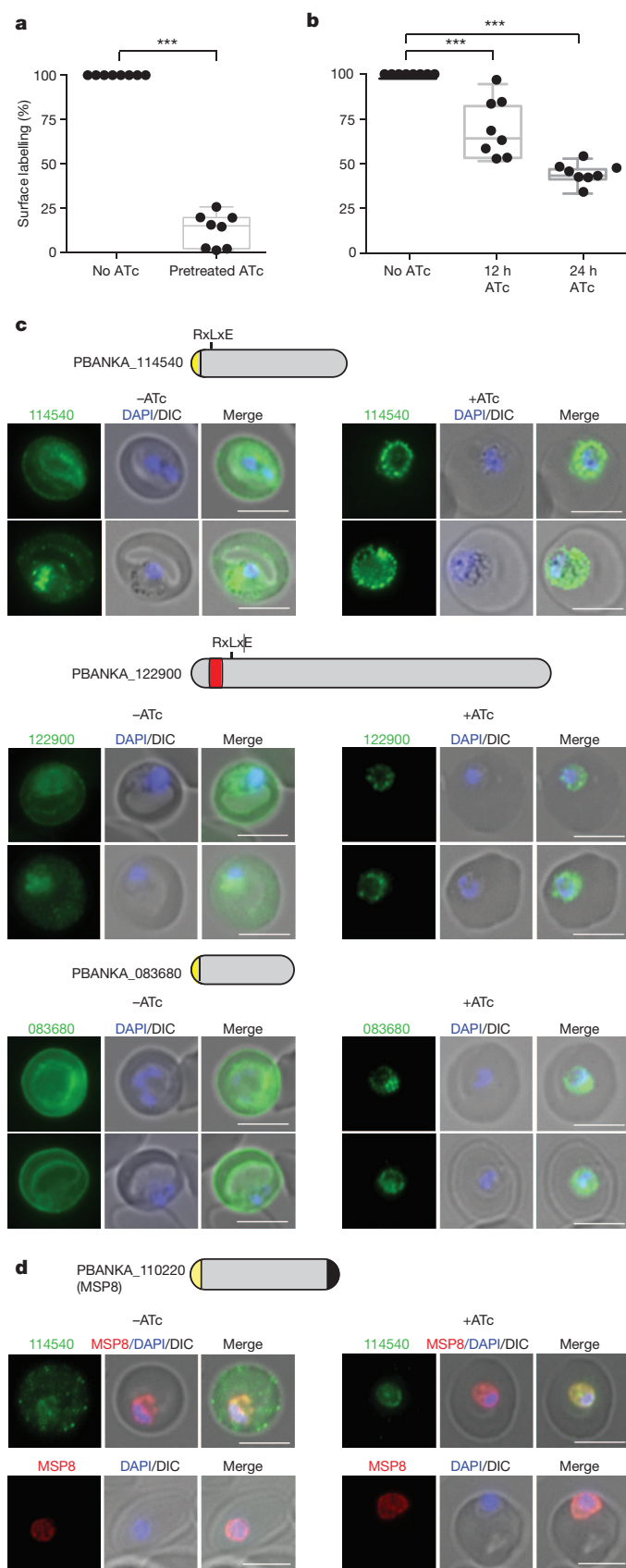


Figure 2 | Knockdown of *P. berghei* HSP101 blocks export of PEXEL and PNEP proteins. **a**, Surface labelling of parasite antigens on Pbi101 KD parasites harvested between days 1 and 2 post infection from mice pretreated with ATc was substantially decreased compared with infected erythrocytes not exposed to ATc as measured by FACS ($n = 8$; error bars represent s.e.m.; $***P < 0.001$, using unpaired t -test). Boxes and whiskers delineate all data points, with whiskers indicating minimum and maximum values. **b**, Surface labelling of parasite antigens on asynchronous Pbi101 KD parasites grown to high parasitaemia and then treated with ATc for either 12 or 24 h ($n = 8$; error bars represent s.e.m.). **c**, Representative IFA of 100 Pbi101 KD intraerythrocytic stages, showing that exposure to ATc blocks export of PEXEL (top and middle panels) and PNEP (bottom panel) proteins. Yellow bar in all diagrams, signal sequence; red bar, transmembrane domain; black bar, glycosylphosphatidylinositol anchor. DIC, differential interference contrast. **d**, Expression of MSP8 is not affected by ATc. Scale bars, 5 μ m.

protein (KAHRP)²³; a double transmembrane and PEXEL-containing protein, Hyp8 (refs 3, 24); and a PNEP, skeleton binding protein 1 (SBP1)²⁵. Both Hyp8 and SBP1 localize to Maurer's clefts, membranous structures that reside in the erythrocyte cytosol. RESA was a significant inclusion because it is exported in very early ring stages, within about 1 h after invasion²⁶ and before the growth inhibitory effect of glucosamine, controlling for any non-specific effect of PTEX150 knockdown on export.

We used two concentrations of glucosamine: the sublethal level of 0.15 mM, to minimize any indirect effect of growth arrest on export, and a high dose of 2.5 mM. Because both RESA and KAHRP are found throughout the cytosol during the ring stages, we quantified their export

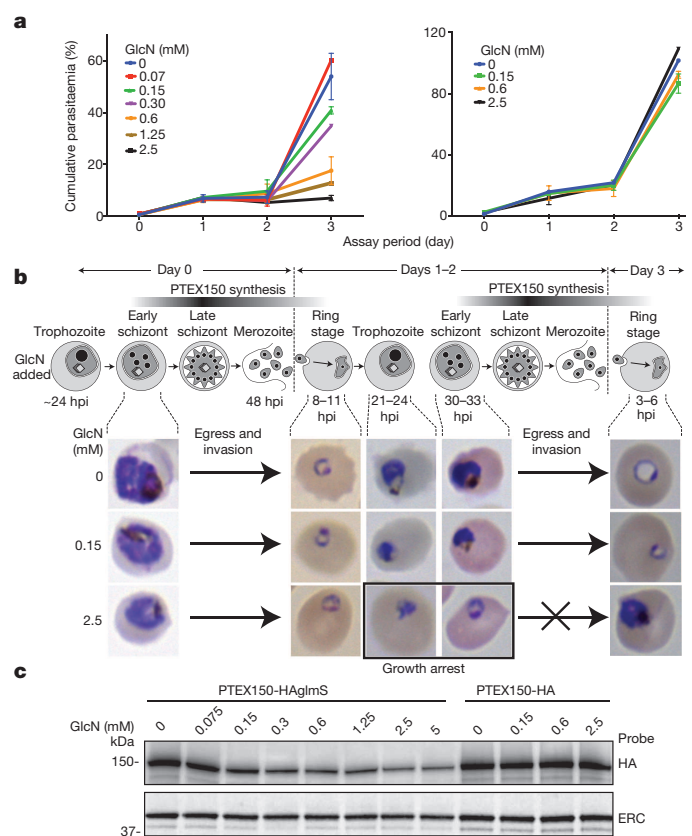


Figure 3 | Generation of a PTEX150 knockdown line in *P. falciparum*. **a**, PTEX150-HAGlmS parasites (left), but not the control PTEX150-HA parasites (right), fail to proliferate when treated with glucosamine (GlcN) at a concentration of 0.6 mM or higher in the previous cycle ($n = 2$). **b**, Giemsa-stained *P. falciparum* cells, showing arrest of growth in the pigmented trophozoite stage (21–33 h post invasion (hpi)) in 2.5 mM GlcN added to the previous cycle. **c**, Western blot analysis: similarly treated PTEX150-HAGlmS parasites, but not the control PTEX150-HA parasites, show an up to 80% decrease in PTEX150 protein levels ($n = 2$). HA, haemagglutinin epitope tag.

With this knockdown approach we used quantitative IFA to examine the export of four different exported proteins: an early expressed PEXEL protein, ring-infected erythrocyte surface antigen (RESA)²²; a 'soluble' PEXEL-containing protein, *Pl. falciparum* knob-associated histidine-rich

by measuring the mean fluorescence intensity in the erythrocyte cytosol at different times after invasion, excluding the region occupied by the parasite (as denoted by staining with 4',6-diamidino-2-phenylindole (DAPI) and staining for EXP2). Using this approach, a strong blockage in export was seen at both 2.5 mM and 0.15 mM glucosamine in PTEX150-HAglms parasites (Fig. 4a, b and Extended Data Fig. 7). This defect was not seen without glucosamine or in the control PTEX150-HA line (Fig. 4a, b). For SBP1 and Hyp8, Maurer's clefts in the erythrocyte cytosol were counted using an automated quantitative microscopic approach. Again, specific knockdown of PTEX150 led to a strong defect in export of these proteins (Fig. 4c, d and Extended Data Figs 8 and 9), both of which seemed to be blocked at the parasitophorous vacuole. To further control for non-specific effects on vesicular trafficking, we examined the localization of MSP8 under PTEX150 knockdown conditions. No effect on either MSP8 expression or localization to the parasite membrane was observed (Fig. 4e).

Finally, we used pooled VAR2CSA reactive immune serum and found that surface PfEMP1 expression was markedly reduced as a result of

PTEX150-specific knockdown, even at relatively low levels of glucosamine (Fig. 4f). Consistent with this, we also demonstrated a smaller proportion of total PfEMP1 expressed on the surface of PTEX150 knock-down parasites by using a cytoadherence assay for chondroitin sulphate A and a trypsin sensitivity assay (Fig. 4g and Extended Data Fig. 6c). Although PfEMP1 requires many PEXEL and PNEP proteins for its trafficking to the infected erythrocyte surface^{7,27}, the fact that under knock-down conditions it remained trapped in the parasite and was not present in the host cell cytosol or at the Maurer's clefts is consistent with the direct trafficking of PfEMP1 by PTEX (Extended Data Fig. 6d).

Thus, using a number of different approaches to specifically knock down the expression of PTEX components we have shown that all parasite protein classes destined for the erythrocyte cytosol were prevented from crossing the parasitophorous vacuole membrane. This is direct evidence that the role of the PTEX molecular machine is to translocate proteins across this membrane (summarized in Fig. 1a). As a nexus for all proteins destined for the host cell cytosol, and directly or indirectly for parasite proteins expressed on the infected erythrocyte surface, this

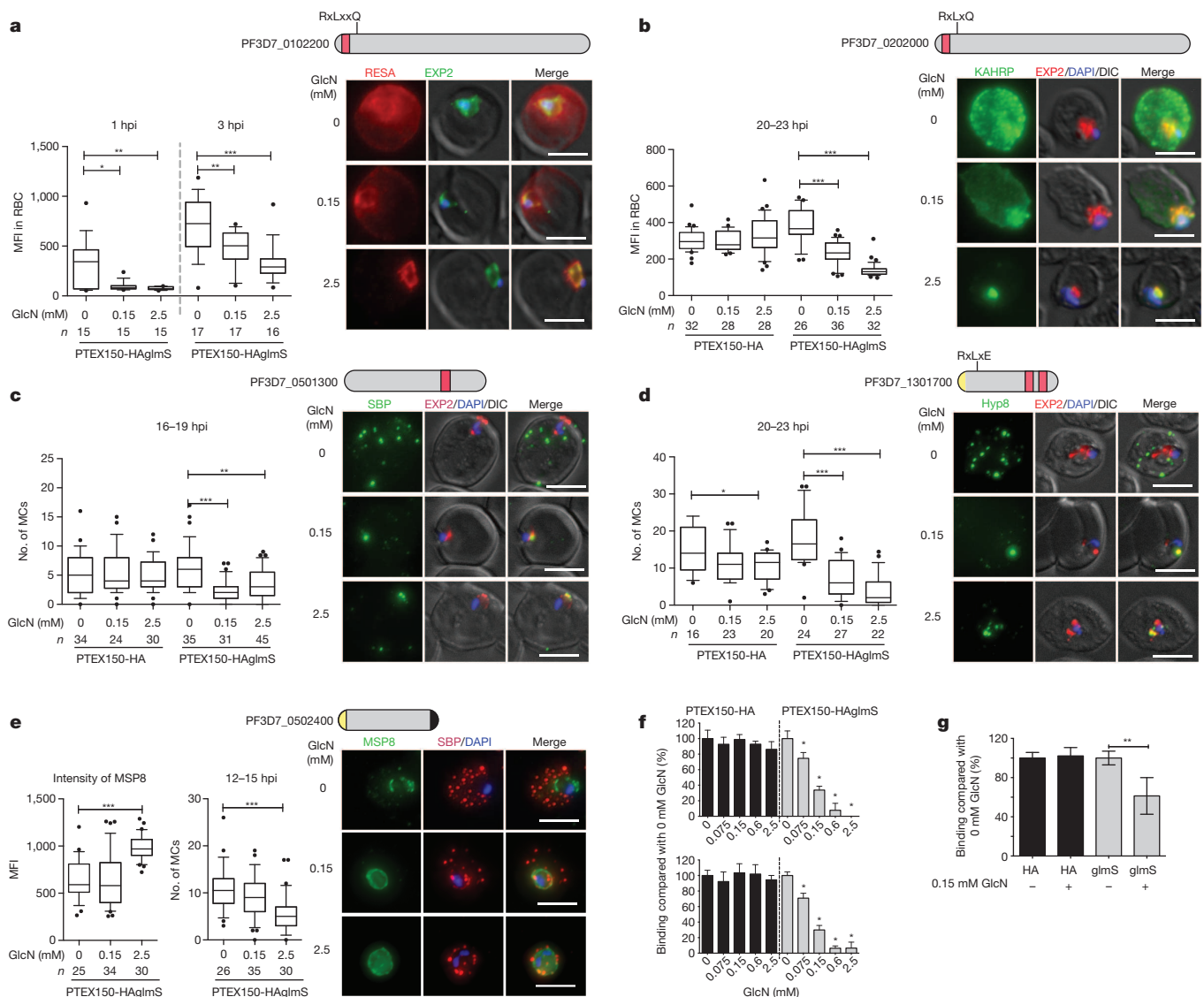


Figure 4 | PTEX150 knockdown blocks protein export in *P. falciparum*. **a–e**, IFAs (right) and graphs (left) showing a decrease in the export of RESA (**a**) and KAHRP (**b**) (mean fluorescence intensity, MFI) and of SBP1 (**c**) and Hyp8 (**d**) (Maurer's clefts, MCs) ($n = 12–47$ cells for each antibody or GlcN concentration) but similar levels of MSP8 (**e**) ($n = 15–35$) after treatment with GlcN. Boxes and whiskers delineate 25th–75th and 10th–90th centiles, respectively. Colours of bars in diagrams as in Fig. 2. Scale bars, 5 μm . **f**, Flow

cytometry analysis showing decreased export of VAR2CSA onto the erythrocyte surface after treatment with GlcN ($n = 3$), at 24–28 h after invasion (top) and at 32–36 h after invasion (bottom). **g**, Cytoadherence of PTEX150-HAglms to chondroitin sulphate A ($n = 2$). Bars represent means \pm s.d. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ as determined by unpaired t -test, with the exception of cytoadherence studies, in which a Mann–Whitney test was used.

complex becomes an attractive target for drug development. Consistent with its potential as a potent drug target, we show here that even modest knockdown of PTEX components had a strong inhibitory effect on *Plasmodium* growth *in vitro*. *In vivo*, the inhibitory effect is likely to be even stronger because the processes involved in pathogenesis, such as those designed to avoid splenic clearance^{7,28,29}, will also be disrupted by compounds inhibiting PTEX function.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 February; accepted 30 May 2014.

Published online 16 July 2014.

- Marti, M., Good, R. T., Rug, M., Knuepfer, E. & Cowman, A. F. Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science* **306**, 1930–1933 (2004).
- Hiller, N. L. *et al.* A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science* **306**, 1934–1937 (2004).
- Sargeant, T. J. *et al.* Lineage-specific expansion of proteins exported to erythrocytes in malaria parasites. *Genome Biol.* **7**, R12 (2006).
- van Ooij, C. *et al.* The malaria secretome: from algorithms to essential function in blood stage infection. *PLoS Pathog.* **4**, e1000084 (2008).
- Boddey, J. A. *et al.* Role of plasmepsin V in export of diverse protein families from the *Plasmodium falciparum* exportome. *Traffic* **14**, 532–550 (2013).
- Heiber, A. *et al.* Identification of new PNEPs indicates a substantial non-PEXEL exportome and underpins common features in *Plasmodium falciparum* protein export. *PLoS Pathog.* **9**, e1003546 (2013).
- Maier, A. G. *et al.* Exported proteins required for virulence and rigidity of *Plasmodium falciparum*-infected human erythrocytes. *Cell* **134**, 48–61 (2008).
- Boddey, J. A. & Cowman, A. F. *Plasmodium* nesting: remaking the erythrocyte from the inside out. *Annu. Rev. Microbiol.* **67**, 243–269 (2013).
- Elsworth, B., Crabb, B. S. & Gilson, P. R. Protein export in malaria parasites: an update. *Cell. Microbiol.* **16**, 355–363 (2014).
- de Koning-Ward, T. F. *et al.* A newly discovered protein export machine in malaria parasites. *Nature* **459**, 945–949 (2009).
- Riglar, D. T. *et al.* Spatial association with PTEX complexes defines regions for effector export into *Plasmodium falciparum*-infected erythrocytes. *Nature Commun.* **4**, 1415 (2013).
- Matthews, K. *et al.* The *Plasmodium* translocon of exported proteins (PTEX) component thioredoxin-2 is important for maintaining normal blood-stage growth. *Mol. Microbiol.* **89**, 1167–1186 (2013).
- Matz, J. M., Matuschewski, K. & Kooij, T. W. Two putative protein export regulators promote *Plasmodium* blood stage development *in vivo*. *Mol. Biochem. Parasitol.* **191**, 44–52 (2013).
- Bullen, H. E. *et al.* Biosynthesis, localisation and macromolecular arrangement of the *Plasmodium falciparum* translocon of exported proteins; PTEX. *J. Biol. Chem.* **287**, 7871–7884 (2012).
- Pino, P. *et al.* A tetracycline-repressible transactivator system to study essential genes in malaria parasites. *Cell Host Microbe* **12**, 824–834 (2012).
- Haase, S., Hanssen, E., Matthews, K., Kalanon, M. & de Koning-Ward, T. F. The exported protein PbCP1 localises to cleft-like structures in the rodent malaria parasite *Plasmodium berghei*. *PLoS ONE* **8**, e61482 (2013).
- Pasini, E. M. *et al.* Proteomic and genetic analysis demonstrate that *Plasmodium berghei* blood stages export a large and diverse repertoire of proteins. *Mol. Cell. Proteomics* **12**, 426–448 (2012).
- Drew, D. R., Sanders, P. & Crabb, B. S. *Plasmodium falciparum* merozoite surface protein 8 is a ring-stage membrane protein that localizes to the parasitophorous vacuole of infected erythrocytes. *Infect. Immun.* **73**, 3912–3922 (2005).
- de Koning-Ward, T. F., Drew, D. R., Chesson, J. M., Beeson, J. & Crabb, B. S. Truncation of *Plasmodium berghei* merozoite surface protein 8 does not affect *in vivo* blood-stage development. *Mol. Biochem. Parasitol.* **159**, 69–72 (2008).
- Prommana, P. *et al.* Inducible knockdown of *Plasmodium* gene expression using the glmS ribozyme. *PLoS ONE* **8**, e73783 (2013).
- Salanti, A. *et al.* Evidence for the involvement of VAR2CSA in pregnancy-associated malaria. *J. Exp. Med.* **200**, 1197–1203 (2004).
- Culvenor, J. G., Day, K. P. & Anders, R. F. *Plasmodium falciparum* ring-infected erythrocyte surface antigen is released from merozoite dense granules after erythrocyte invasion. *Infect. Immun.* **59**, 1183–1187 (1991).
- Pologe, L. G. & Ravetch, J. V. A chromosomal rearrangement in a *P. falciparum* histidine-rich protein gene is associated with the knobless phenotype. *Nature* **322**, 474–477 (1986).
- Sleeb, B. E. *et al.* Inhibition of Plasmepsin V activity demonstrates its essential role in protein export, PfEMP1 display and survival of malaria parasites. *PLoS Biol.* **12**, e1001897 (2014).
- Blisnick, T. *et al.* Pfsbp1, a Maurer's cleft *Plasmodium falciparum* protein, is associated with the erythrocyte skeleton. *Mol. Biochem. Parasitol.* **111**, 107–121 (2000).
- Riglar, D. T. *et al.* Super-resolution dissection of coordinated events during malaria parasite invasion of the human erythrocyte. *Cell Host Microbe* **9**, 9–20 (2011).
- Cooke, B. M. *et al.* A Maurer's cleft-associated protein is essential for expression of the major malaria virulence antigen on the surface of infected red blood cells. *J. Cell Biol.* **172**, 899–908 (2006).
- Cooke, B. M., Rogerson, S. J., Brown, G. V. & Coppel, R. L. Adhesion of malaria-infected red blood cells to chondroitin sulfate A under flow conditions. *Blood* **88**, 4040–4044 (1996).
- Crabb, B. S. *et al.* Targeted gene disruption shows that knobs enable malaria-infected red cells to cytoadhere under physiological shear stress. *Cell* **89**, 287–296 (1997).

Acknowledgements We thank T. Templeton, B. Franke-Fayard, C. Janse, A. Cowman, J. Boddey, B. Cooke, M. Duffy, L. Tilley, R. Anders, F. Fowkes, A. McLean and D. Bursac for reagents and/or other assistance with aspects of this study; D. Stanisic, F. Baiwog and I. Mueller for contributions to clinical studies of pregnant women; and P. Siba. We also thank the Australian Red Cross Blood Bank for the provision of human blood and serum. This work was supported by grants from the National Health and Medical Research Council (NHMRC) of Australia (1021560, 1025665 and 637406) and the Victorian State Government Operational Infrastructure Support Scheme. T.F.d.K.-W. is an NHMRC Career Development Fellow, and J.G.B. is a NHMRC Senior Research Fellow. B.E. and K.M. are the recipients of Australian Postgraduate Awards.

Author Contributions B.E., K.M., P.R.G. and T.F.d.K.-W. designed, performed and interpreted much of the experimental work. B.S.C. designed and interpreted the work and, along with P.R.G. and T.F.d.K.-W., wrote the manuscript. C.Q.N., M.K., S.C.C., P.R.S., S.A.C. and N.A.C. performed experiments and provided intellectual insight into aspects of this study. P.J.S., P.P., J.C., M.F.A., J.G.B. and S.J.R. provided reagents and intellectual input into study design. All authors commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.F.d.K.-W. (taniad@deakin.edu.au), B.S.C. (crabb@burnet.edu.au) or P.R.G. (gilson@burnet.edu.au).

METHODS

Infection of mice with *P. berghei* parasites. Female Balb/c mice (6–8 weeks of age) were randomized into groups and infected intraperitoneally with 10^7 parasitized erythrocytes. Parasitaemias were determined by Giemsa-stained blood smears; at least 1,000 erythrocytes were counted. All experiments involving mice were performed in strict accordance with the recommendations of the Australian Government and the NHMRC Australian code of practice for the care and use of animals for scientific purposes. Protocols were approved by the Deakin University Animal Welfare Committee (approval no. AWC A97/10).

Plasmid constructs. The construct pTg-ranTRAD4-iHSP101 was used to generate the *P. berghei* i101 KD. This plasmid was based on pPRF-TRAD4-Tet07-HAPRF-hDHR¹⁵ but modified to include a BsiWI restriction site downstream of the profilin coding sequence and a BssHII restriction enzyme site between the profilin 5' untranslated region and the TRAD4 sequence. This enabled cloning of the first 1.7 kilobases (kb) of the HSP101 coding sequence (PbANKA_09312; amplified with DO390F, 5'-caccctgcagATGGTACGGAACATTGCTAAAAATT-3', and DO414R, 5'-gtatctgacgcatggCTATAACTCTTGGTTTACCCG-3', the latter containing an internal NcoI site) into the PstI and BsiWI cloning sites and 0.85 kb of the HSP101 5' untranslated region (amplified using oligonucleotides DO392F, 5'-gtaccatggCGTACGGTATGCAATTGCTCTTAATGCATTTGC-3', and DO394R, 5'-tatgcgagcTTTCTACTAAATTTATAGTAAATATAGATATA-3') into the NcoI and BssHII sites. Before transfection, DNA was linearized with NcoI.

pPTEX150-HA-glmS was produced by excision of the strep-tag from pPTEX150-HA/Str3' (ref. 10) and the introduction of a stop codon followed by a 3' glmS sequence.

Transfection. The reference clone 15cyl from the *P. berghei* ANKA strain was used to generate *P. berghei* transgenic parasites. Transfection of schizont-stage parasites with 1 µg of linearized DNA construct was performed with the Nucleofector electroporation device (Amaxa), using previously described protocols³⁰; stable transfectants were selected by adding pyrimethamine at a final concentration of 0.07 mg ml⁻¹ to the drinking water of mice. For the Pb101 KD parasites, PCR and Southern blot analysis (performed as outlined below) of the pyrimethamine-resistant population obtained after transfection revealed that the parasites were already clonal. Nevertheless, limiting-dilution cloning was performed on this population, and both the original homogeneous population and cloned line were used for phenotypic analysis.

For *P. falciparum*, 100 µg of pPTEX150-HA and pPTEX150-HAglmS were used to transfect CS2 parasites that had recently been selected for chondroitin sulphate A (CSA) binding³¹. Transfected parasites were selected with 2.5 nM WR99210 and were cycled on and off the drug to select for integration into the *ptex150* locus.

Nucleic acid analysis. The genotype of the *P. berghei* i101 KD was confirmed by Southern blot analysis of genomic DNA isolated from infected rodent blood. Nucleic acid probes were synthesized using the DIG PCR Probe Synthesis kit (Roche); detection was performed with the DIG Luminescent Detection kit (Roche) in accordance with the manufacturer's protocol. PCR was also used to confirm integration at the 5' and 3' ends and the purity of the population using a combination of the following oligonucleotides: a, 5'-TTATAGTTTGAACACCAAGGACG-3'; b, 5'-GCCTTCGATACCGACTTCATTGAG-3'; c, 5'-CTTTCGATACCGTCGACCTCGAG-3'; d, 5'-TTTTGCTTAATGGCTCGAAAA-3'. To detect PTEX transcripts in *P. berghei* ANKA parasites by RT-PCR, RNA was extracted from blood-stage parasites, using the NucleoSpin RNA II Kit (Macherey-Nagel) followed by treatment with DNaseI (Invitrogen). cDNA was then made using the Omniscript RT Kit (Qiagen) in accordance with the manufacturer's manual. cDNA (or genomic DNA as a control) was used in PCR reactions using the oligonucleotides to detect *hsp101* (DO390, 5'-ATGGTACGGAACATTGCTAAAAATT-3'; DO391, 5'-CCAAATTGTTCAATGTTTAATCCAG-3'), *rap1* (DO186, 5'-GATTATTCTGTGCAATTTAACAT-3'; DO187, 5'-GAAGGTAATCATTTTTTGTGG-3') and *ptex150* (MK28, 5'-AATGACCAGCCAATTGTTCC-3'; MK29, 5'-TGCATCTTTGCCTTCTTCCT-3').

Correct integration of the plasmids into the *ptex150* locus (Pf3D7_1436300) was confirmed by PCR on DNA template isolated from the parasite lines using primers A (5'-CGTTGTAAATTTCTAAATATGCTGATAATTCC-3'), B (5'-TTCTTTTATTTTCTTTCTTTAGCTCTCCATTGT-3') and C (5'-CCGGGACGTCGTACGGGTATGCTG-3').

Treatment with ATc. For the *in vivo* exposure of parasites to ATc, mice were randomized into groups of three for each experiment and then given drinking water containing 0.2 mg ml⁻¹ ATc (Sigma) made in 5% sucrose. ATc was either administered to mice from 24 h before infection or, in other experiments where indicated, parasites were grown to a higher parasitaemia (~5–10%) in mice in the absence of ATc, and then exposed to ATc for designated durations. A minimum of 1,000 erythrocytes were counted to determine the parasitaemia. For *in vitro* treatment, parasites harvested from mice at ring stage were grown until schizont stage at 36.5 °C (~16 h) in RPMI 1640 medium containing L-glutamine (Life Technologies) supplemented with 25 mM HEPES, 0.2% bicarbonate, 25% fetal bovine serum and 1 µg ml⁻¹ ATc (or vehicle as a control).

Treatment with glucosamine. *P. falciparum* CS2 PTEX150-HA and CS2 PTEX150-HAglmS parasites were synchronized with 5% sorbitol; at 24 h after invasion, 1 M glucosamine (Sigma) was added to various final concentrations (0.075–2.5 mM) as well as a 0 mM glucosamine control. For microscopy, later in the same cell cycle, heparin sulphate (Sigma) was added to prevent new invasions. The heparin was washed out and added back 3 h later to give the parasites a 3 h invasion window. At times corresponding to 4–7, 8–11, 12–15, 16–19 and 20–23 h after invasion, thin blood smears were made from all cultures and smears were allowed to dry in air. Slides were stored at -20 °C until needed. To make highly synchronous young ring-stage parasites, merozoites were prepared as described³² and allowed to invade for 10 min before heparin sulphate was added to inhibit new invasions. At times corresponding to 1 and 3 h after invasion, thin blood smears were made from all cultures and treatments were allowed to dry in air before being stored frozen until needed.

Western blot analysis. *P. berghei* ring-stage parasites administered to mice pre-exposed to ATc or vehicle control were harvested 29 h later and lysed with 0.09% saponin. Equal volumes of parasite material were fractionated in 8% acrylamide Bis-Tris gels and blotted onto 0.45 µm poly(vinylidene difluoride) membrane (Millipore). The membranes were blocked in 3% BSA in PBS and probed with rabbit anti-HSP101 (1:200 dilution), rabbit anti-EXP2 (1:200) or rabbit anti-MSP8 (1:1,000). After washing, the membranes were probed with horseradish peroxidase-conjugated secondary antibodies, and detection was performed with SuperSignal enhanced chemiluminescence (Thermo Fischer Scientific). Quantification of signal strength was performed with NIH ImageJ version 1.47d. Western blots were performed on samples made from two independent experiments and were analysed twice.

P. falciparum parasites were treated with various concentrations of glucosamine when about halfway through their cell cycle. At mid ring stage in the next cycle (~12 h after invasion) the parasites were harvested and were freeze-thawed once to break open the erythrocyte compartment and release the haemoglobin, which was then washed out with PBS. Equal amounts of parasite proteins were fractionated in 4–12% acrylamide Bis-Tris gels (Novex Life Technologies) and blotted onto nitrocellulose membrane. The membranes were blocked in 1% casein in PBS and were probed with the following primary antibodies: chicken anti-HA epitope (1:1,000; Abcam), rabbit anti-ERC (1:500), rabbit anti-HSP70-1 (1:500), rabbit anti-GAPDH (1:2,000), rabbit anti-HSP101 (1:500) and a mouse anti-RESA monoclonal antibody (mAb 1812; 1 µg ml⁻¹). After washing, the membranes were probed with fluorescent secondary antibodies and detected with a Li-Cor Odyssey FC scanner. Band densities were measured using the scanner's software. Four separate expression assays were performed and showed similar trends. The densitometry data presented here are from one assay analysed twice by western blotting.

Growth assays. *P. berghei* schizonts obtained from overnight parasite cultures grown in 150 ml of complete RPMI were burst by mechanical action, using a fine needle syringe. Viable merozoites, purified by filtration through a 0.2 µm filter, were resuspended in fresh medium and combined 1:1 with fresh erythrocytes. Invasion was allowed to proceed for 30 min at 37 °C with vigorous shaking. Cells were either maintained in culture *in vitro* or intravenously injected into naive mice to establish synchronous mouse infections.

For *P. falciparum* assays, glucosamine was added to a final concentration of 0.075, 0.15, 0.3, 0.6, 1.25 or 2.5 mM, with 0 mM glucosamine serving as the negative control. Thin blood smears of the cultures were air-dried before being stained with Giemsa. A minimum of 2,000 erythrocytes were counted to determine the parasitaemia of the culture.

Immunofluorescence analysis of *P. berghei*. Erythrocytes infected with *P. berghei* parasites were fixed for 30 min with 4% paraformaldehyde and 0.0075–0.015% glutaraldehyde in PBS. After washing, cells were permeabilized for 10 min with either 0.25% Triton X-100 or 0.5% Triton X-100 where indicated, and then washed a further three times, before blocking for 30 min with 1% BSA/PBS (Sigma). Cells were labelled for 1 h at 20–22 °C or 16 h at 4 °C with anti-EXP2 (1:300), anti-ACP (1:300), anti-PbANKA_114540 (1:150), anti-PbANKA_122900 (1:150), anti-PbANKA_083680 (1:150) or anti-MSP8 (1:1,000) (all raised in rabbits), washed and then sequentially labelled for 1 h with goat anti-rabbit AlexaFluor 488/568 secondary antibody (1:2,000; Life Technologies). Cells were mounted in Vectashield containing the nuclear stain DAPI (VectorLabs). For each experiment, at least 100 parasitized erythrocytes were counted on two occasions by different researchers; where quantification was performed, samples were blinded. For co-labelling experiments, anti-PbANKA_114540 (1:500) and goat anti-rabbit secondary AlexaFluor 488 antibody (1:300–500) were added sequentially for 1 h, followed by either anti-EXP2 (1:300), anti-ACP (1:300) or no antibody for 16 h, with a final incubation for 1 h with goat anti-rabbit AlexaFluor 568 secondary antibody (1:2,000). For localization of PbANKA_114540 and anti-MSP8, anti-MSP8 (1:1,000) and goat anti-rabbit secondary AlexaFluor 568 antibody (1:300) were added sequentially for 1 h, followed by anti-PbANKA_114540 or no antibody for 16 h, with a final incubation for 1 h with goat anti-rabbit AlexaFluor 488 secondary antibody (1:2,000). Images were acquired independently by at least

two researchers with an Olympus IX70 microscope and processed using NIH ImageJ version 1.47d or Adobe CS6 Photoshop.

P. falciparum IFA, image scoring and statistics. For *P. falciparum* IFA, blood smears of parasites were thawed, then fixed in ice-cold 100% methanol for 5 min before being air-dried. Three time courses were used: a short one for RESA in which parasites were synchronous within a 10-min window, a mid-range course for SBP, Hyp8, KAHRP and MSP8 in which parasites were within a 3 h window, and a long course for PfEMP1 where the window was 5 h. The parasites were rehydrated and blocked for 1 h in 3% BSA (Sigma) in PBS. The cells were then probed with anti-EXP2 ($3 \mu\text{g ml}^{-1}$), anti-SBP (1:200) and anti-RESA ($1 \mu\text{g ml}^{-1}$) mouse monoclonal antibodies and rabbit serum for SBP (1:200), KAHRP (1:1,000), Hyp8 (1:200), PfEMP1 ATS (1:500) or MSP8 (1:1,000) diluted in 3% BSA in PBS. After being washed three times in PBS, the cells were probed with goat anti-mouse AlexaFluor 568 (1:2,000; Invitrogen) and goat anti-rabbit AlexaFluor 488 (1:2,000; Invitrogen) in 3% BSA in PBS for 1 h. After being washed three times in PBS, the cells were mounted in Vectashield with DAPI. To avoid potential bias, the parasites were selected for imaging solely on the basis of their DAPI (nuclear) staining and then imaged in the green fluorescent protein, Texas Red, ultraviolet and DIC channels on a Zeiss Axio Observer microscope. For each time point the same exposure times were used, to produce consistent fluorescence intensities. After imaging 23 Z sections at $0.28 \mu\text{m}$ apart for each cell, a Z-projection was made and used to score the degree of protein export.

Before the images were scored in FIJI v.1.48, their file names were de-identified of parasite line information. To score the mean level of MSP8 expression, in the whole infected erythrocytes the cell area was selected and the mean fluorescence intensity was obtained by means of the 'Measure' function. To score the degree of KAHRP and RESA export, the circumference of the infected erythrocyte was first traced around, followed by the parasite (denoted by EXP2 and DAPI staining) to exclude it from subsequent export analysis. The mean fluorescence intensity of KAHRP and RESA was then quantified as above. The labelling of SBP and Hyp8 in the infected erythrocyte cytosol were similarly traced and the punctate Maurer's clefts were counted, using the 'Find Maxima' function set to a noise tolerance of 200 and 'Point Selection' output type. The mean fluorescence intensities for KAHRP-labelled cells and the number of Maurer's clefts for SBP and Hyp8 were graphed in GraphPad Prism. Unpaired *t*-tests using parametric distribution were performed to measure differences between glucosamine-treated and untreated parasites and were assumed to be significant when $P < 0.05$. A Mann-Whitney test was used for the RESA data.

Flow cytometry. For analysis of *P. berghei* surface antigens using fluorescence-activated cell sorting (FACS), 20 μl of blood collected from the tail vein of *P. berghei*-infected mice was washed briefly in RPMI and then blocked for 1 h in 1% casein in RPMI. Erythrocytes were then incubated for 1 h with serum harvested from either *P. berghei* semi-immune or non-immune (pre-bleed) mice, generated as described³³ and diluted 1:20 in blocking solution. After three washes with block solution, cells were incubated for 1 h with goat anti-mouse IgG AlexaFluor 647 (1:2000; Invitrogen), washed a further three times and then incubated for 5 min in Sybr safe (Invitrogen) diluted 1:2000 in blocking solution. A further three washing steps were performed, after which the cell preparation was analysed with a FACS Canto II machine (BD Biosciences). All incubations were performed at $20\text{--}22^\circ\text{C}$.

P. falciparum cultures were synchronized with sorbitol and then CS2 PTEX150-HA and CS2 PTEX150-HAGlms parasites were cultured in the presence of glucosamine at indicated concentrations at 24–30 h after invasion. Heparin sulphate ($100 \mu\text{g ml}^{-1}$; Sigma) was also added to prevent new invasion events³⁴. Parasites incubated with no glucosamine served as controls. Heparin was removed to permit invasion; 4 h later, parasites were synchronized with 5% sorbitol to remove trophozoites and unruptured schizonts, resulting in an invasion window of 4 h (ref. 32). Samples were taken at 24–28, 32–36 and 38–42 h after invasion, and the surface expression of the PfEMP1 variant VAR2CSA was measured²⁵.

To test for surface VAR2CSA expression on these parasites, serum samples were collected from multigravid women that attended antenatal care at Alexishafen Health Centre, Madang Province, Papua New Guinea, which is endemic for *P. falciparum* malaria^{35,36}. Serum samples were screened by ELISA for reactivity against VAR2CSA-DBL5 recombinant protein^{37,38}, and a pool of five high responder samples was made; reactivity of serum IgG in this pool to the surface of intact CS2-infected erythrocytes was confirmed using flow cytometry. A pool of serum from malaria-naïve individuals from Melbourne, Australia, was used as a negative control. Serum IgG binding to the surface of infected erythrocytes was measured by flow cytometry as described previously^{39,40}. In brief, samples taken at various time points were washed in PBS with 0.1% casein, resuspended at 0.2% haematocrit, and incubated sequentially with test serum diluted 1:40 in 0.1% casein in PBS, then with polyclonal rabbit anti-human IgG (1:100) in 0.1% casein, and lastly with $10 \mu\text{g ml}^{-1}$ goat anti-rabbit AlexaFluor 488 and $10 \mu\text{g ml}^{-1}$ ethidium bromide in 0.1% casein. All incubations were for 30 min

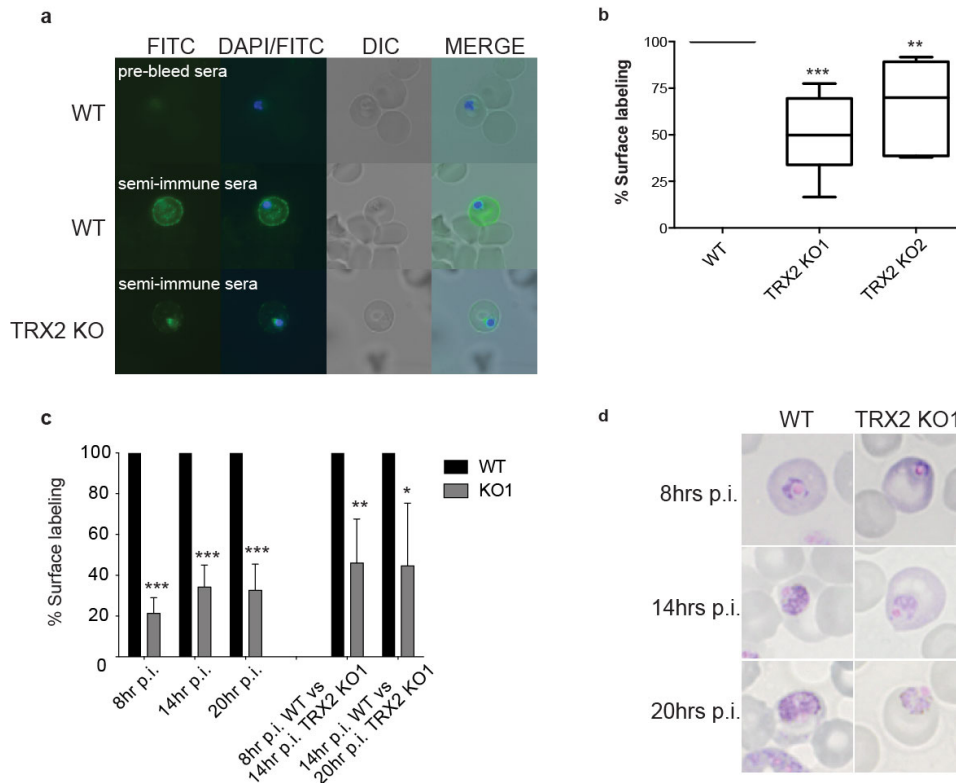
at $20\text{--}22^\circ\text{C}$, with three washes in 0.1% casein between incubations. Cells were resuspended in PBS, and data were acquired with a FACS Canto II machine and analysed with FlowLogic software (eBioscience). After gating for erythrocytes, serum IgG binding for each sample was expressed as the geometric mean fluorescence intensity (MFI) for trophozoites after subtracting the MFI of uninfected erythrocytes. Samples were considered as positive for IgG binding when the MFI was more than three s.d. above the mean binding seen with malaria-naïve control sera. Ethics approval for human studies was provided by the Alfred Hospital Human Research and Ethics Committee, and the Medical Research Advisory Committee of Papua New Guinea. All participants gave written informed consent.

Cleavage of surface proteins with trypsin. *P. falciparum* parasites were cultured in RPMI-HEPES supplemented with 10% human serum. Glucosamine was added at 0 and 0.3 mM 24 h after invasion CS2 PTEX150-HA and CS2 PTEX150-HAGlms parasites. Midway through the following cell cycle, the parasites were purified by magnetic separation and were treated with RPMI-HEPES with 5% sucrose, with or without 1 mg ml^{-1} trypsin, for 1 h at 37°C (ref. 41). Soybean trypsin inhibitor was added to stop the reaction, and the parasites were solubilized in 1% Triton X-100 on ice for 20 min and centrifuged at $18,000\text{g}$ for 5 min. The pellet fraction containing PfEMP1 was solubilized by sonication in 2% SDS at $20\text{--}22^\circ\text{C}$ and equal amounts were separated by electrophoresis on 3–8% Tris-acetate polyacrylamide (Novex Life Technologies) and transferred to a nitrocellulose membrane. The membrane was probed with monoclonal anti-PfEMP1 acidic terminal segment (ATS) (1B/98 8AB-19-18 at $6 \mu\text{g ml}^{-1}$) and goat anti-mouse antibody IRDye 800 Conjugated (Rockland) at 1:5,000 and imaged as above.

Cytoadherence assays. Adhesion of CS2 PTEX150-HAGlms parasites and control PTEX150-HA to the receptor CSA was tested after treatment with 0.15 mM glucosamine (+) in the previous cycle. After a 30 min incubation of gelatin-enriched parasites with immobilized CSA (from bovine trachea, 20 mg ml^{-1}), unbound cells were washed and bound parasites were fixed with 2% glutaraldehyde in PBS, stained with Giemsa and counted by microscopy⁴². The number of cells adhered per square millimetre for each condition (tested in triplicate) was normalized to its (–) glucosamine control. The graph shows results from two independent experiments. Error bars represent s.d.; $*P < 0.005$ (Mann-Whitney test).

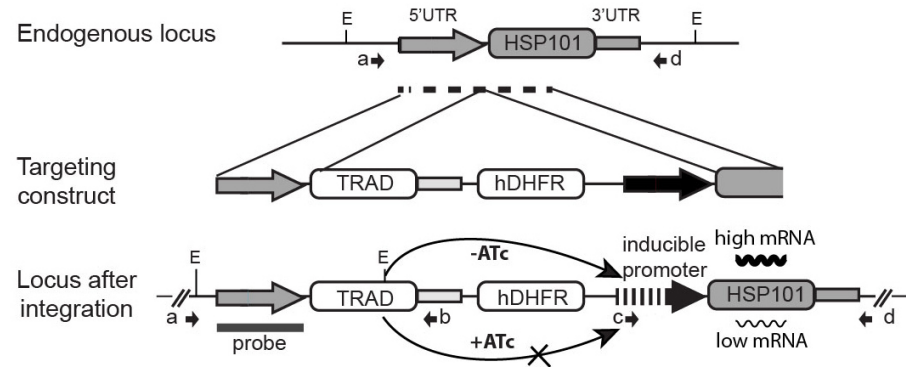
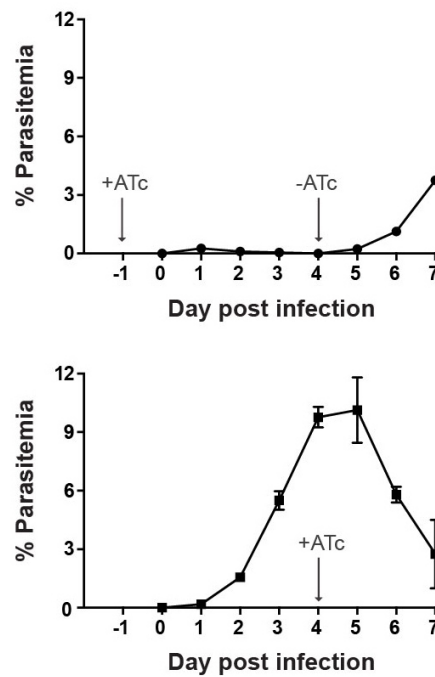
Statistics. All graphs and data generated in this study were analysed using GraphPad Prism 6.0b Software (MacKiev). Unpaired *t*-tests using parametric distribution were performed to measure differences between untreated and treated (ATc or glucosamine) parasites. $P < 0.05$ was considered significant.

30. Janse, C. J., Ramesar, J. & Waters, A. P. High-efficiency transfection and drug selection of genetically transformed blood stages of the rodent malaria parasite *Plasmodium berghei*. *Nature Protocols* **1**, 346–356 (2006).
31. Rogerson, S. J., Chaiyaroj, S. C., Ng, K., Reeder, J. C. & Brown, G. V. Chondroitin sulfate A is a cell surface receptor for *Plasmodium falciparum*-infected erythrocytes. *J. Exp. Med.* **182**, 15–20 (1995).
32. Boyle, M. J. *et al.* Isolation of viable *Plasmodium falciparum* merozoites to define erythrocyte invasion events and advance vaccine and drug development. *Proc. Natl Acad. Sci. USA* **107**, 14378–14383 (2010).
33. de Koning-Ward, T. F. *et al.* A new rodent model to assess blood stage immunity to the *Plasmodium falciparum* antigen merozoite surface protein 1₁₉ reveals a protective role for invasion inhibitory antibodies. *J. Exp. Med.* **198**, 869–875 (2003).
34. Boyle, M. J., Richards, J. S., Gilson, P. R., Chai, W. & Beeson, J. G. Interactions with heparin-like molecules during erythrocyte invasion by *Plasmodium falciparum* merozoites. *Blood* **115**, 4559–4568 (2010).
35. Umbers, A. J. *et al.* Placental malaria-associated inflammation disturbs the insulin-like growth factor axis of fetal growth regulation. *J. Infect. Dis.* **203**, 561–569 (2011).
36. Beeson, J. G. *et al.* Antigenic differences and conservation among placental *Plasmodium falciparum*-infected erythrocytes and acquisition of variant-specific and cross-reactive antibodies. *J. Infect. Dis.* **193**, 721–730 (2006).
37. Avril, M. *et al.* Immunization with VAR2CSA-DBL5 recombinant protein elicits broadly cross-reactive antibodies to placental *Plasmodium falciparum*-infected erythrocytes. *Infect. Immun.* **78**, 2248–2256 (2010).
38. Hommel, M. *et al.* Evaluation of the antigenic diversity of placenta-binding *Plasmodium falciparum* variants and the antibody repertoire among pregnant women. *Infect. Immun.* **78**, 1963–1978 (2010).
39. Beeson, J. G. *et al.* Antibodies to variant surface antigens of *Plasmodium falciparum*-infected erythrocytes and adhesion inhibitory antibodies are associated with placental malaria and have overlapping and distinct targets. *J. Infect. Dis.* **189**, 540–551 (2004).
40. Chan, J. A. *et al.* Targets of antibodies against *Plasmodium falciparum*-infected erythrocytes in malaria immunity. *J. Clin. Invest.* **122**, 3227–3238 (2012).
41. Maier, A. G. *et al.* Skeleton-binding protein 1 functions at the parasitophorous vacuole membrane to traffic PfEMP1 to the *Plasmodium falciparum*-infected erythrocyte surface. *Blood* **109**, 1289–1297 (2007).
42. Beeson, J. G. *et al.* *Plasmodium falciparum* isolates from infected pregnant women and children are associated with distinct adhesive and antigenic properties. *J. Infect. Dis.* **180**, 464–472 (1999).



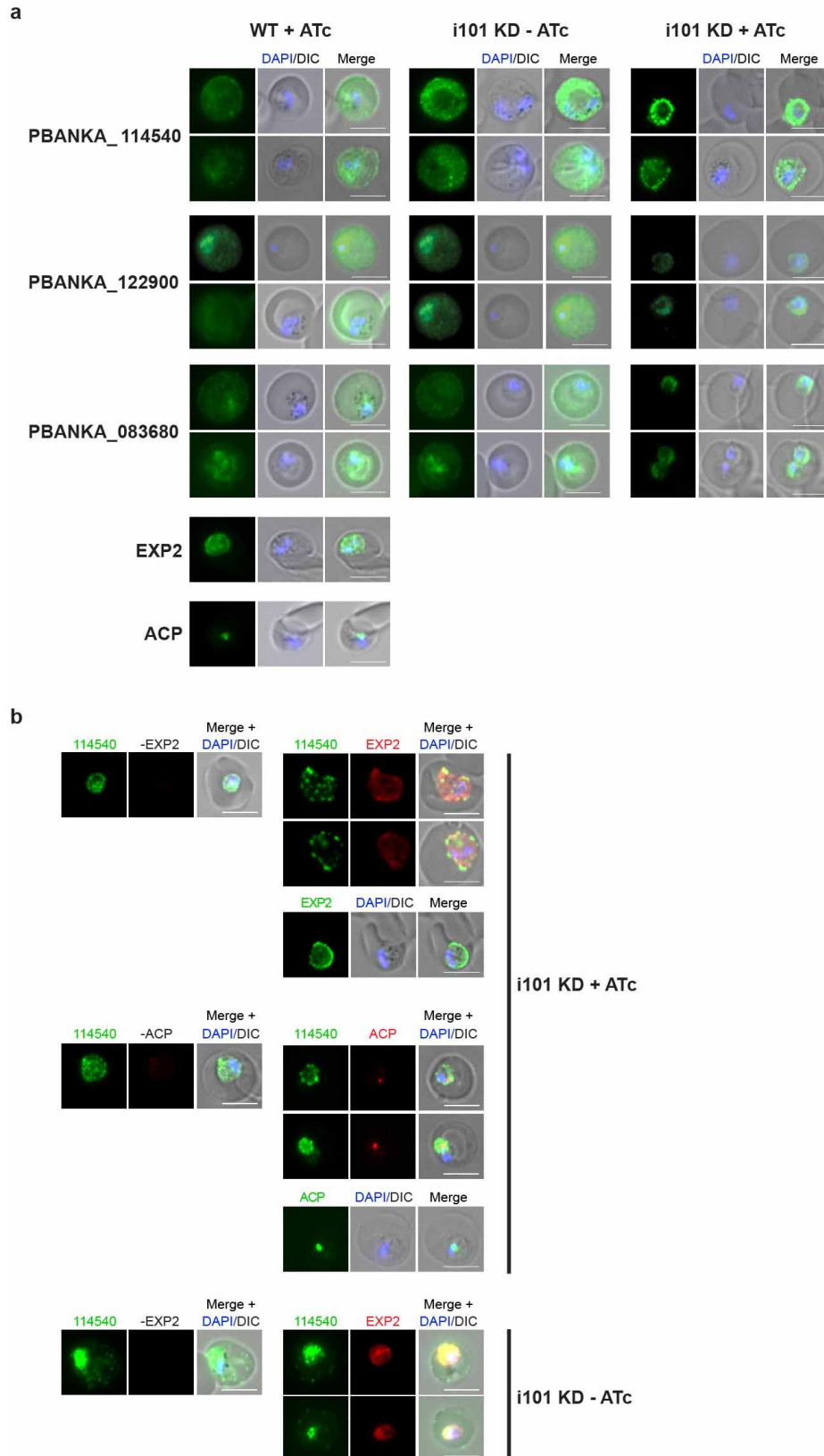
Extended Data Figure 1 | Disruption of *P. berghei* TRX2 leads to reduced protein export. **a**, IFA of fixed infected erythrocytes using *P. berghei* semi-immune sera reveals TRX2 knockout parasites (TRX2 KO) show reduced surface labelling compared with wild-type *P. berghei* ANKA parasites (WT), indicative of a reduction in expression of parasite antigens on the surface of erythrocytes infected with the TRX2 KO. Pre-bleed sera were used as a negative control. **b**, Quantitative FACS analysis of erythrocytes harvested from asynchronously infected mice ($n = 6$) show that two independent clonal populations of TRX2 KO parasites exhibit significantly reduced levels of surface labelling with *P. berghei* semi-immune sera compared with wild-type parasites ($*P < 0.05$; $**P < 0.01$; $***P < 0.001$, unpaired t -test). **c**, As **b**, except that

synchronous mouse infections were initiated by injecting purified merozoites into the tail veins of mice, and surface labelling of infected erythrocytes with semi-immune sera was performed at time points relative to when the wild-type line reinvaded erythrocytes for the second cycle (left); p.i., post invasion. Even taking into consideration that disruption of TRX2 leads to slower growth by about 6 h, the surface labelling of TRX2 KO parasites at a stage of growth comparable to that of wild-type parasites is also significantly reduced (right) ($n = 3$ independent experiments). **d**, Giemsa smears showing the stages of parasite development at time points relative to when wild-type parasites had invaded erythrocytes.

a**b**

Extended Data Figure 2 | Generation of a HSP101 knockdown line in *P. berghei*. **a**, Schematic representation used to construct Pbi101 KD parasites. PCR primers used to detect 5' integration (a/b), 3' integration (c/d) and wild-type locus (a/d) are indicated. E, EcoRI. **b**, Representative experiments ($n = 3$) showing parasitaemias in mice that were (upper panel) or were not

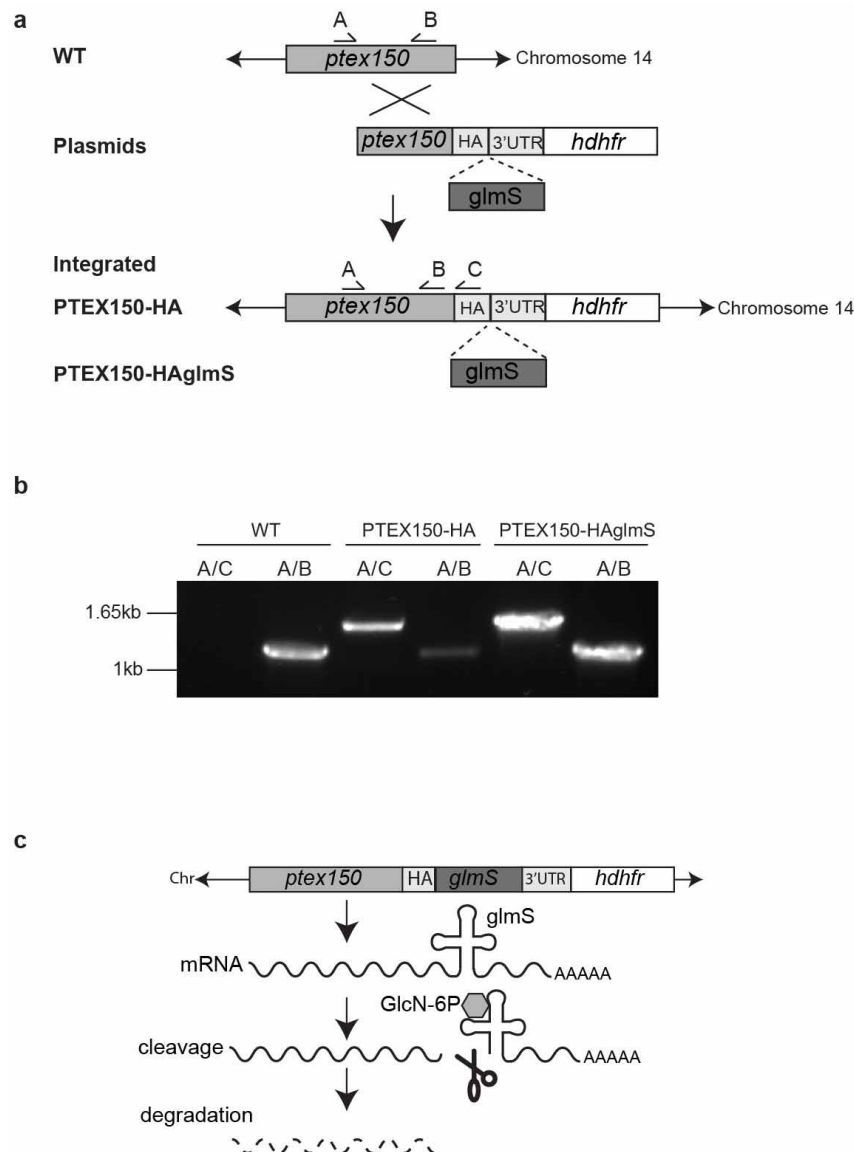
(lower panel) pre-exposed to ATc in their drinking water before infection. At day 4 after infection, the treatment regimens in both experiments were switched. Error bars show s.e.m. for three mice per condition performed in parallel.



Extended Data Figure 3 | Knockdown of HSP101 blocks protein export.

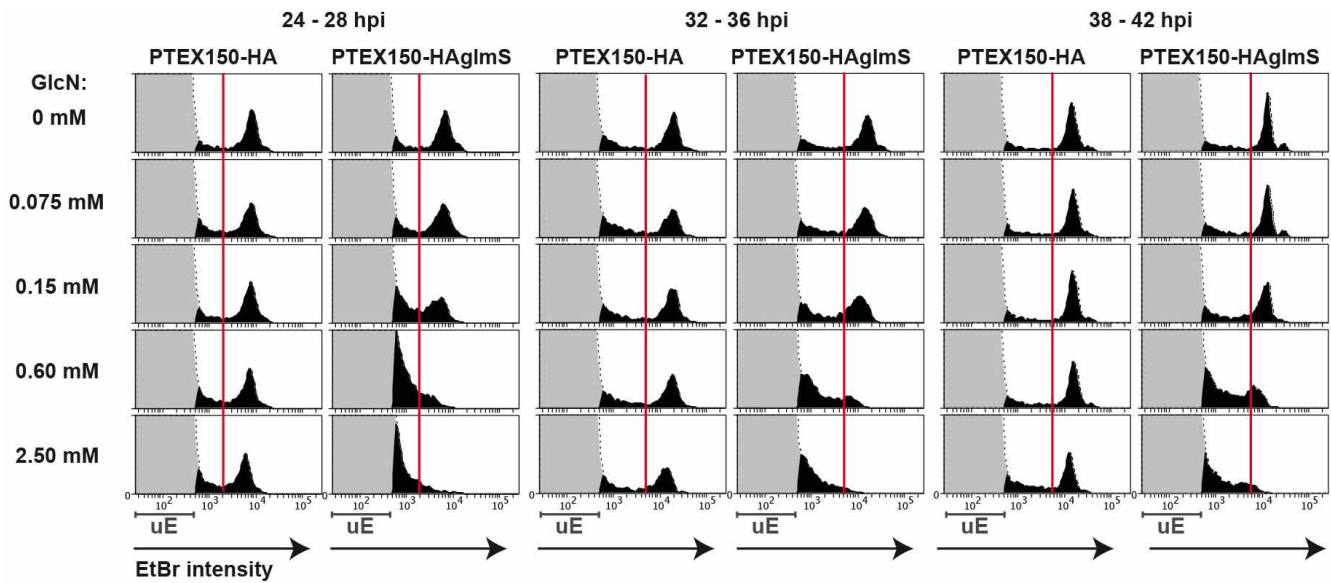
a, Representative IFA of intraerythrocytic stages showing that export of three different *P. berghei* proteins across the parasitophorous vacuole membrane is blocked when Pbi101 KD, but not wild-type, is exposed to ATc. Samples were harvested at the times indicated by the asterisks in Fig. 1d and e. **b**, IFAs show that correct localization of EXP2 and ACP is unaffected in Pbi101 KD parasites treated with ATc (right panels). In these samples, cells were

permeabilized after fixation with 0.5% Triton X-100. Because the PbANKA_114540, EXP2 and ACP antibodies were all raised in rabbits, sequential labelling with anti-PbANKA_114540, anti-rabbit AlexaFluor488, anti-EXP2 or anti-ACP, and anti-rabbit AlexaFluor568 had to be performed. Control IFAs were therefore performed in which anti-EXP2 or anti-ACP were omitted (left panels).



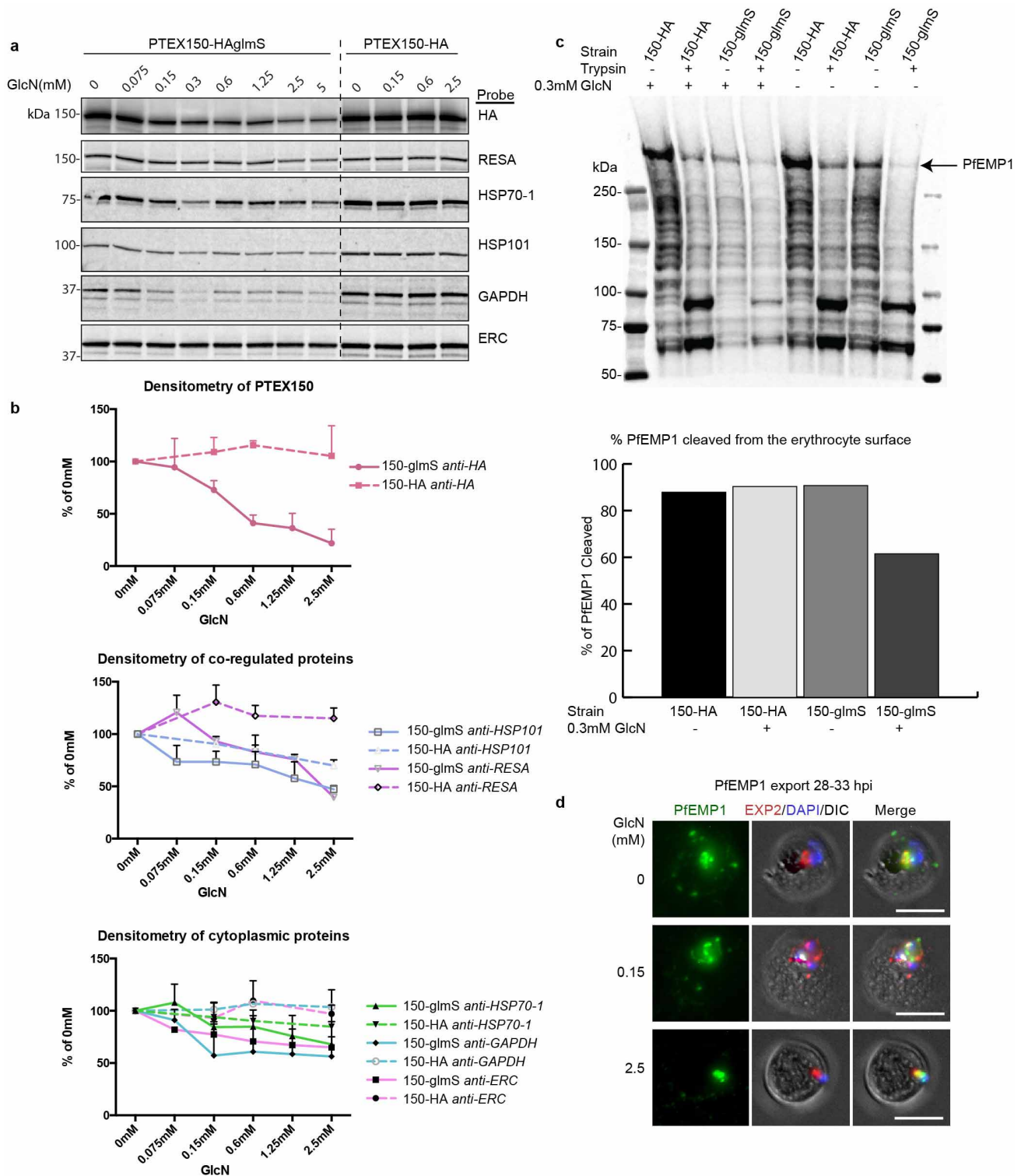
Extended Data Figure 4 | Diagnostic PCR analysis shows the *ptex150* gene has been appended with a HA tag in the PTEX150-HA parasites and a HAglmS tag in the PTEX150-HAglmS parasites. **a**, Diagram of the targeted genetic crossovers and binding sites of the PCR primers. **b**, Using the indicated primer combinations, correct 3' - recombination has occurred in the

PTEX150-HA and PTEX-HAglmS parasites using primers A/C, with a band specific to the integrated locus (1.7 kb) only observed in HA-tagged parasite lines. **c**, Diagram showing how the *glmS* ribozyme after glucosamine binding is stimulated to cleave its mRNA, resulting in message destabilization and a decrease in protein levels.



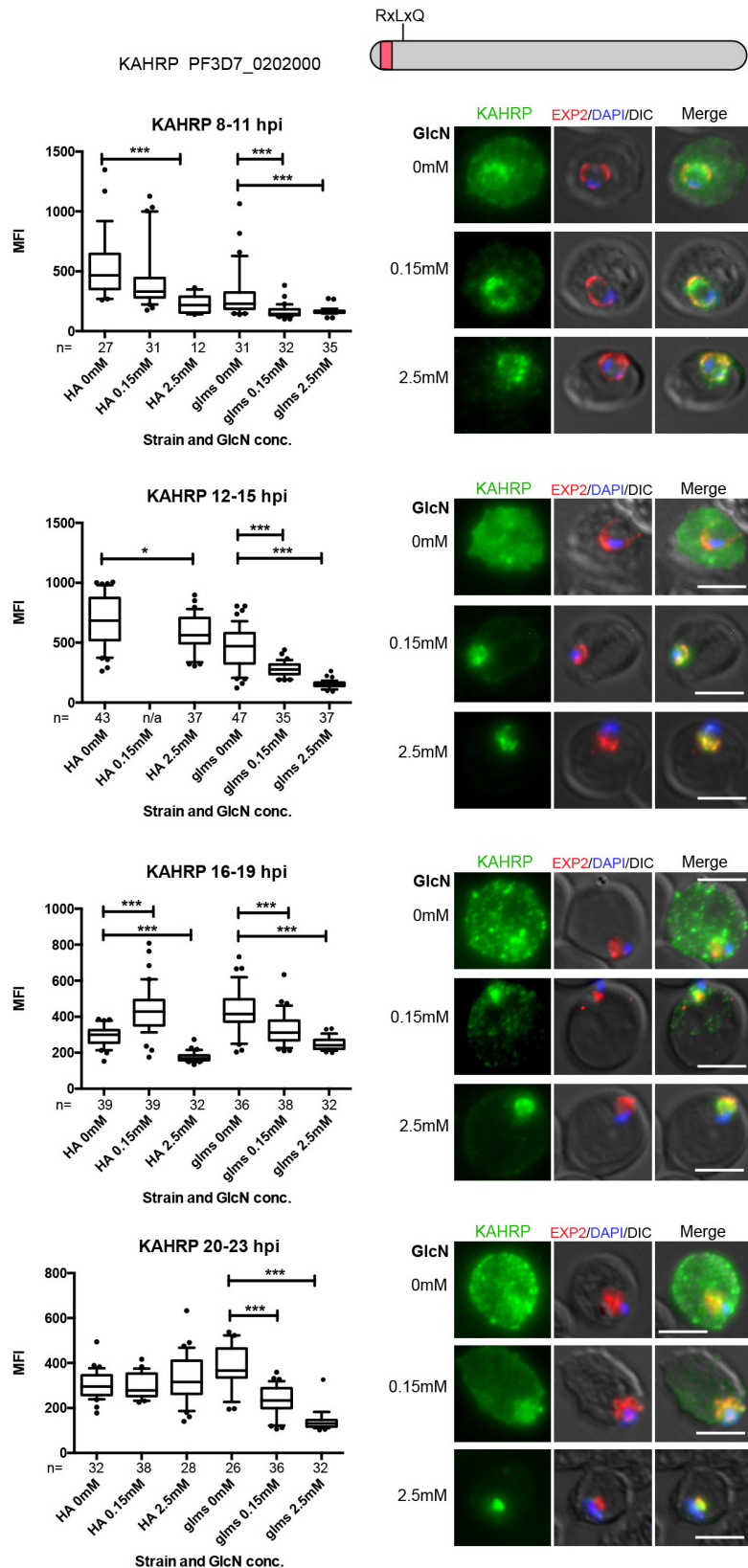
Extended Data Figure 5 | Growth assays of PTEX150-HA and PTEX150-HAgImS parasites show that growth of the latter declines substantially after treatment with glucosamine (GlcN). CS2 PTEX150-HA and CS2 PTEX150-HA-gImS parasites were treated with different concentrations of glucosamine from 24–30 h after invasion (hpi) and then allowed to invade fresh erythrocytes for 4 h. At the times after invasion indicated, the cells were stained with ethidium bromide to measure DNA content as a marker for parasite

growth. Representative histograms show the levels of ethidium bromide intensity (*x* axis) and cell number (*y* axis). Infected and uninfected erythrocytes (uE) are shown as black and grey, respectively. Those parasites to the right of the red line are the strongly staining trophozoites; those to the left are the younger, weakly staining ring stages. Assays were performed at least three times independently.



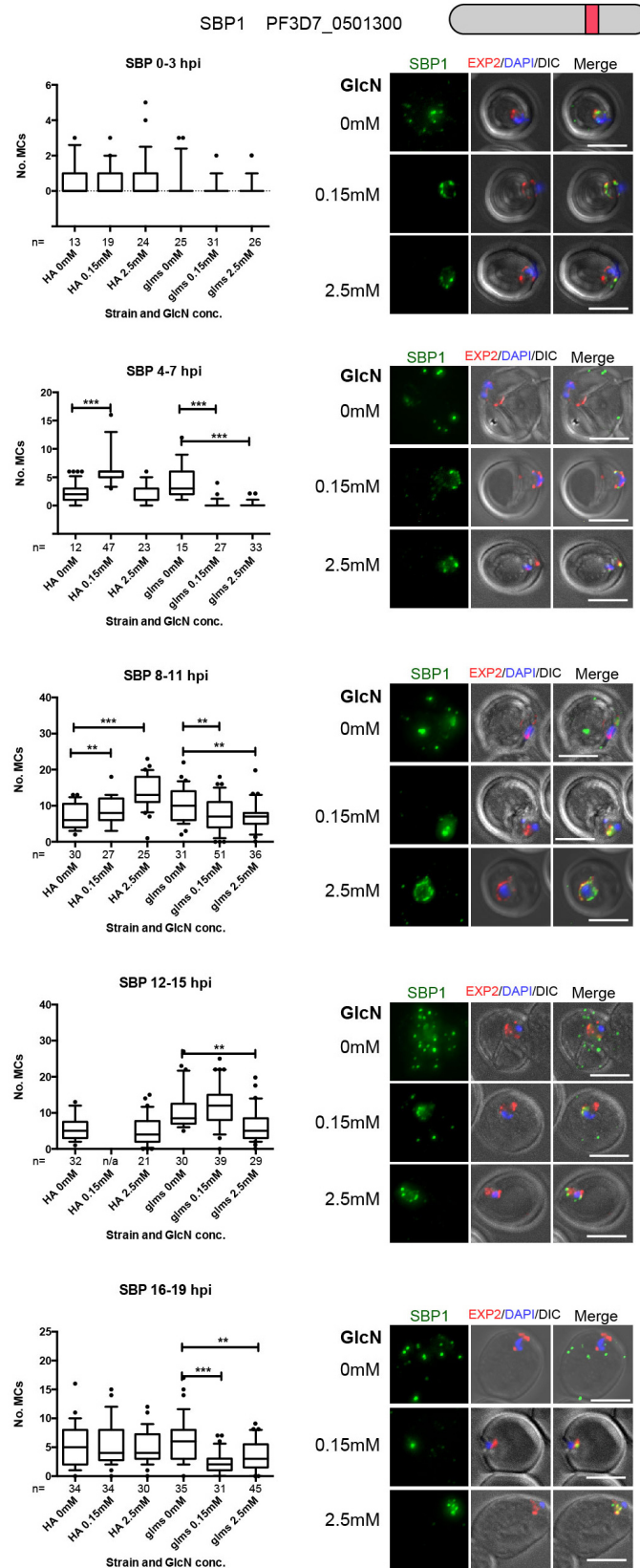
Extended Data Figure 6 | PTEX150-HAglmS protein levels are markedly reduced on induction of the glmS ribozyme with GlcN. **a**, Western blots of PTEX150-HAglmS and control PTEX150-HA mid-ring-stage parasites (~12 hpi) probed with the antibodies indicated on the right. GlcN was added at the concentration indicated above the blots, halfway through the previous cell cycle. **b**, Western blots were performed in duplicate and densitometry of the bands has been graphed showing the mean \pm s.d. relative to no GlcN. Top: PTEX150 levels in the PTEX150-HAglmS (150-glmS) decrease with increasing concentrations of GlcN to a minimum ~17% of the level without GlcN. The levels of PTEX150 in the control PTEX150-HA (150-HA) parasites does not decrease in GlcN. Middle and bottom: the levels of co-regulated HSP101 and RESA proteins and cytoplasmic constitutive HSP70-1, GAPDH and ERC

proteins also decline in the PTEX150-HAglmS parasites after treatment with GlcN to about 50–60%, indicative of slowed growth due to loss of PTEX150 function. **c**, Western blot of infected erythrocytes treated with trypsin to cleave off surface-exposed PfEMP1. The blot has been probed with a monoclonal antibody against the intracellular C-terminal tail of PfEMP1, and the densitometry of the 350 kDa VAR2CSA band (arrow) has been compared between trypsin-treated and untreated infected erythrocytes to calculate the percentage cleaved in the presence or absence of 0.3 mM GlcN. **d**, IFAs of PTEX150-HAglmS probed for PfEMP1 and EXP2 after treatment with GlcN indicate a decrease in the export of PfEMP1-containing structures to the periphery of the infected erythrocyte. Scale bar, 5 μ m.



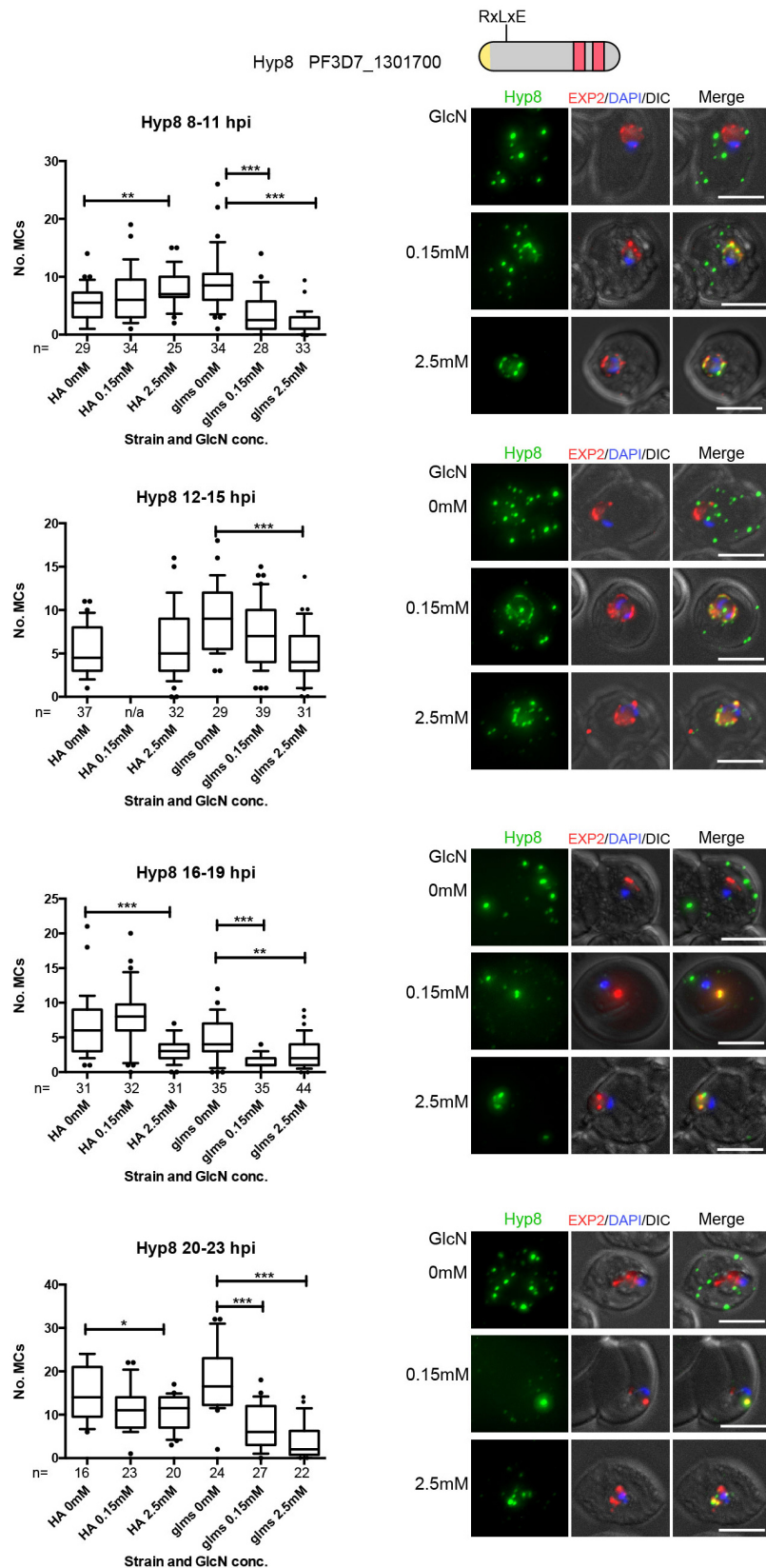
Extended Data Figure 7 | Export of KAHRP in PTEX150-HAglms (glms) is decreased after treatment with GlcN. The mean fluorescence intensity (MFI) of the erythrocyte compartment in infected erythrocytes stained with rabbit anti-KAHRP always declines after the addition of GlcN halfway through the previous cell cycle. In comparison, treatment with GlcN does not consistently decrease KAHRP export in the control PTEX150-HA (HA) parasites; the variation is possibly due to inconsistencies in sample preparation. In the graphs,

the boxes and whiskers delineate the 25–75th and 10–90th centiles, respectively. Outlying data points are shown as dots. Significances: * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ by unpaired t -test. The number of cells (n) counted is indicated below the graph. Example immunofluorescence images of only PTEX150-HAglms are shown. The regions occupied by the parasite are indicated by staining with DAPI and staining for EXP2. Scale bar, 5 μ m.



Extended Data Figure 8 | Export of SBP1 in PTEX150-HAGlmS (glms) is decreased after treatment with GlcN. The number of punctate Maurer's clefts (MCs) present in the erythrocyte compartment in infected erythrocytes stained with rabbit anti-SBP1 nearly always declines after the addition of GlcN halfway through the previous cell cycle. In comparison, treatment with GlcN does not consistently decrease SBP1 export in the control PTEX150-HA (HA) parasites; the variation is possibly due to inconsistencies in sample

preparation. In the graphs, the boxes and whiskers delineate the 25–75th and 10–90th centiles, respectively. Outlying data points are shown as dots. Significances: * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ by unpaired t -test. The number of cells (n) counted is indicated below the graph. Example immunofluorescence images of only PTEX150-HAGlmS are shown. The regions occupied by the parasite are indicated by staining with DAPI and staining for EXP2. Scale bar, 5 μ m.



Extended Data Figure 9 | Export of Hyp8 in PTEX150-HAGlms (glms) is reduced following glucosamine treatment. The number of punctate Maurer's Clefts (MCs) present in the erythrocyte compartment in infected erythrocytes stained with rabbit anti-Hyp8 nearly always declines following addition of GlcN half way through the previous cell cycle. In comparison, GlcN treatment does not consistently reduce Hyp8 export in the control PTEX150-HA (HA) parasites and the variation is possibly due to inconsistencies in sample

preparation. In the graphs, the boxes and whiskers border the 25–75th and 10–90th percentiles, respectively. Outlying data points are shown as dots. Significances: * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$ by unpaired t -test. The number of cells (n) counted is indicated below the graph. Example immunofluorescence images of only PTEX150-HAGlms are shown. The regions occupied by the parasite are indicated by staining with DAPI and staining for EXP2. Scale bar, 5 μ m.

PTEX component HSP101 mediates export of diverse malaria effectors into host erythrocytes

Josh R. Beck^{1*}, Vasant Muralidharan^{2,3*†}, Anna Oksman^{1,2,3} & Daniel E. Goldberg^{1,2,3}

To mediate its survival and virulence, the malaria parasite *Plasmodium falciparum* exports hundreds of proteins into the host erythrocyte¹. To enter the host cell, exported proteins must cross the parasitophorous vacuolar membrane (PVM) within which the parasite resides, but the mechanism remains unclear. A putative *Plasmodium* translocon of exported proteins (PTEX) has been suggested to be involved for at least one class of exported proteins; however, direct functional evidence for this has been elusive^{2–4}. Here we show that export across the PVM requires heat shock protein 101 (HSP101), a ClpB-like AAA + ATPase component of PTEX. Using a chaperone auto-inhibition strategy, we achieved rapid, reversible ablation of HSP101 function, resulting in a nearly complete block in export with substrates accumulating in the vacuole in both asexual and sexual parasites. Surprisingly, this block extended to all classes of exported proteins, revealing HSP101-dependent translocation across the PVM as a convergent step in the multi-pathway export process. Under export-blocked conditions, association between HSP101 and other components of the PTEX complex was lost, indicating that the integrity of the complex is required for efficient protein export. Our results demonstrate an essential and universal role for HSP101 in protein export and provide strong evidence for PTEX function in protein translocation into the host cell.

PTEX is a protein complex found in the PVM, where a translocon responsible for export into the host erythrocyte is expected². The timing of its synthesis is appropriate for a putative translocon and it has a component (EXP2) with weak homology to bacterial haemolysins, which could form a membrane-spanning channel. PTEX contains two additional core components: the novel protein PTEX150 and HSP101, a member of the Clp/HSP100 family of AAA + ATPases. To directly explore the role of HSP101/PTEX in protein export, we used a conditional auto-inhibition approach by fusing a dihydrofolate reductase (DHFR)-based destabilization domain (DDD) to the endogenous HSP101 carboxy terminus (Fig. 1a). Although this fusion strategy was originally developed to mediate conditional protein degradation via the proteasome^{5,6}, we have found that the destabilized tag can conditionally interfere with chaperone function without protein degradation⁷. Following transfection, two independent clones were isolated which had undergone the intended recombination event (Extended Data Fig. 1).

When the DDD-stabilizing small molecule trimethoprim (TMP) was removed from asynchronous HSP101^{DDD} cultures, a complete block in growth was observed, with parasites accumulating as late ring-stage forms (Fig. 1b, c). Although growth inhibition was TMP-concentration-dependent, this effect was not a result of HSP101^{DDD} degradation (Extended Data Fig. 2a, b). To determine sensitivity to HSP101^{DDD} destabilization across the asexual developmental cycle, TMP was removed at regular intervals in synchronized HSP101^{DDD} parasites. Removal of TMP in early ring-stages resulted in growth arrest. In contrast, when TMP was removed at or after the beginning of the trophozoite stage (18–24 h post-invasion), parasite development and re-invasion proceeded normally, followed by arrest at the subsequent ring stage (Fig. 1d and Extended Data Fig. 2c, d).

Parasite growth was rescued by adding back TMP as late as 48 h after withdrawal, indicating that arrested ring forms remained viable.

As protein export and PTEX expression also occur in parasite sexual stages, we further evaluated the ability of HSP101^{DDD} parasites to form

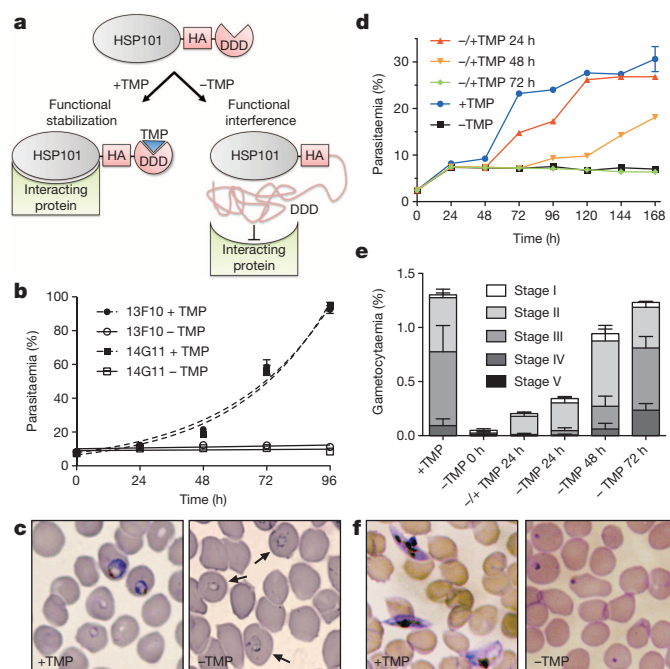


Figure 1 | HSP101 is essential for development of asexual and sexual blood stages. **a**, Auto-inhibition strategy for HSP101^{DDD}. TMP, trimethoprim; HA, haemagglutinin tag; DDD, DHFR destabilization domain. **b**, Growth analysis of asynchronous cultures of the two independent clones 13F10 and 14G11 with or without TMP. Error bars represent s.d. of three technical replicates. Data are representative of three independent experiments. **c**, Giemsa-stained smears of cultures following 48 h with or without TMP. Accumulation of late ring-stage parasites is observed in the absence of TMP (arrows). Images are representative of three independent experiments. **d**, Growth analysis of synchronous 13F10 parasites. TMP was removed at the early trophozoite stage and added back to cultures after 24, 48 or 72 h. Equivalent parasitaemia in all samples at 24 h shows that development through trophozoite and schizont stages, egress and reinvasion were not affected by TMP removal. Error bars as in **b**. Data are representative of three independent experiments. **e**, Analysis of gametocyte formation by 13F10 parasites. TMP was removed from late schizonts following gametocyte induction (0 h) or at subsequent 24 h intervals. In one sample, TMP was removed at 0 h and restored after 24 h (-/+ TMP 24 h). Gametocytaemia of various stages on day nine post-induction is shown. Error bars as in **b**. Data are representative of four independent experiments. **f**, Giemsa-stained smears of gametocyte cultures 9 days post induction. Images are representative of four independent experiments. **d**, **f**, Original magnification $\times 1,000$.

¹Department of Molecular Microbiology, Washington University School of Medicine, St Louis, Missouri 63110, USA. ²Department of Medicine, Washington University School of Medicine, St Louis, Missouri 63110, USA. ³Howard Hughes Medical Institute, Washington University School of Medicine, St Louis, Missouri 63110, USA. [†]Present address: Center for Tropical and Emerging Global Diseases and Department of Cellular Biology, University of Georgia, Athens, Georgia 30602, USA.

*These authors contributed equally to this work.

gametocytes required for malaria transmission^{8,9}. In gametocytogenesis induction experiments, gametocyte development was severely inhibited when TMP was removed during stage I (Fig. 1e, f and Extended Data Fig. 3). Development was partially rescued by re-introducing TMP after 24 h. Consistent with PTEX degradation in later gametocyte stages⁹, gametocyte formation was largely unaffected by later TMP removal (stage II onwards, 48 and 72 h), although a slight delay in development was observed. Collectively, these results demonstrate that HSP101 serves an essential function during asexual and sexual blood-stage development.

To determine the impact of HSP101^{DDD} inhibition on protein export, we examined a panel of proteins representative of the diversity of solubility states, targeting motifs, host-cell destinations and expression timing observed among the *P. falciparum* repertoire of exported effectors. Most exported proteins contain a pentameric *Plasmodium* export element (PEXEL)^{10,11} motif that is cleaved by the aspartic protease plasmepsin V in the parasite endoplasmic reticulum to license these proteins for export^{12,13}. We first examined histidine-rich protein II (HRP2)¹⁴, a soluble PEXEL protein that is exported into the erythrocyte cytosol at the ring stage. To this end, TMP was removed from synchronized HSP101^{DDD} schizonts, which were then allowed to develop for 18–24 h into rings. Remarkably, export of HRP2 into the host cytosol was completely blocked in nearly all infected red blood cells (RBCs), with the protein accumulating in the parasitophorous vacuole (PV) surrounding the parasite (Fig. 2a, b). Similar results were obtained with the soluble, PEXEL-containing protein REX3¹⁵ and the membrane-associated, PEXEL-containing protein KAHRP¹⁶ (Extended Data Fig. 4a, b and g). Among PEXEL-containing proteins, ring-infected erythrocyte surface antigen (RESA) is uniquely stored in secretory organelles of invading merozoites and discharged into the PV along with PTEX immediately following invasion^{3,4,17}. Robust block in RESA export was seen in recently invaded parasites (≤ 1 h post invasion), indicating that, in the absence of TMP, HSP101^{DDD} function is impaired from the time of invasion (Extended Data Fig. 4c, g).

PEXEL-negative exported proteins (PNEPs) are not cleaved by plasmepsin V¹⁸ but seem to contain export signals in their amino termini¹⁹. As PTEX has thus far been implicated only in translocation of PEXEL proteins^{2–4}, we examined the PNEP REX1, a protein associated with

the cytosolic face of parasite-induced structures in the RBC cytosol called Maurer's clefts, which serve as platforms for sorting of exported proteins²⁰. We again observed a block in export with REX1 accumulating in a ring around the parasite where it colocalized with HSP101^{DDD} (Fig. 2c and Extended Data Fig. 4g). A similar block was observed for SBP1²¹ and REX2¹⁵, integral membrane PNEPs exported to the Maurer's clefts (Extended Data Fig. 4d–g). Export was readily restored when TMP was added back to blocked cultures and restoration was largely insensitive to cycloheximide, indicating that previously synthesized HSP101^{DDD} and exported proteins in the PV are sufficient to reactivate export (Extended Data Fig. 5).

As later stages of parasite growth and development are not sensitive to TMP removal (Fig. 1d), we wondered if an export block could be induced after the ring stage. To this end, TMP was removed from synchronous HSP101^{DDD} parasites at the late ring stage and export of trophozoite/schizont-specific proteins was assessed 12–24 h later. We first examined MSRP6, a soluble PNEP peripherally associated with Maurer's clefts²². Parasites lacking TMP displayed a marked block of MSRP6 export whereas SBP1, exported before the block, was found in the erythrocyte (Fig. 2d and Extended Data Fig. 4g). Similar results were seen for the PNEP PfEMP1, a variable surface antigen and key virulence determinant for *P. falciparum* malaria²³ (Fig. 2e and Extended Data Fig. 4g). Importantly, PfEMP1 trafficking to the RBC surface is dependent on several additional exported proteins¹. Thus, although our results indicate delivery of PfEMP1 to the RBC surface is HSP101-dependent, we cannot exclude the possibility that this block is an indirect result of a failure to export other proteins required for PfEMP1 trafficking rather than a direct effect of HSP101 inactivation. Finally, we examined PFA660, a PEXEL-containing HSP40 found in parasite-induced membrane structures in the RBC called J-dots²⁴. Again, washout of TMP efficiently blocked export as visualized by expression of a PFA660–green fluorescent protein (GFP) fusion protein in live parasites (Fig. 2f and Extended Data Fig. 4g). Together, these data show that HSP101 function is necessary for export, but not for parasite development in the trophozoite and schizont stages.

Formation of a normal digestive vacuole containing hemozoin (Fig. 2d–f, DIC panels) indicates that haemoglobin uptake and degradation proceeded normally in these parasites, a process requiring proteases that

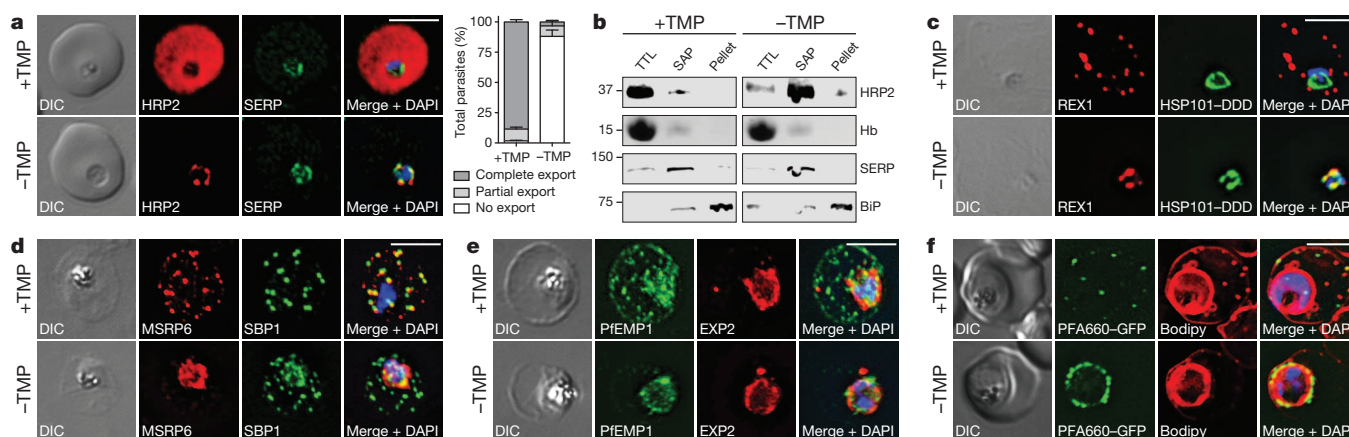


Figure 2 | HSP101 is required for export of PEXEL and PNEP proteins.

a, c, Immunofluorescence assay (IFA) of ring-stage 13F10 parasites with or without TMP. TMP was removed in late schizont stage and parasites were allowed to reinfect and grow 18–24 h before fixation with paraformaldehyde (**a**) or acetone (**c**). **a**, IFA of the exported PEXEL-containing protein HRP2. SERP is a marker for the PV. Export was scored as complete (no HRP2 signal enrichment around the parasite as shown in the +TMP IFA), partial (HRP2 signal within the host cell but also enriched around the parasite) or no export (HRP2 signal only seen around the parasite and not in the host cell, as shown in the –TMP IFA). Error bars represent s.d. of three technical replicates. Data are representative of five independent experiments. DIC, differential interference contrast. **b**, Sequential fractionation of infected ring-stage parasites with or without TMP analysed by western blot. The host cytosol was released with

tetanolysin (TTL) and subsequently the PV contents were released with saponin (SAP). Blocked HRP2 is found in the PV fraction. Haemoglobin (Hb) was detected by Coomassie staining and serves as a control for host cytosol release. SERP serves as a control for PV release. BiP serves as a parasite integrity control. Data are representative of two independent experiments. **c**, IFAs of the PNEP REX1, which colocalizes with HSP101^{DDD} at the PVM in the absence of TMP. **d, e**, IFA of trophozoite-stage 13F10 parasites with or without TMP. TMP was removed in late ring stage and parasites were allowed to develop 12–24 h before fixation with acetone. **f**, Live fluorescence imaging of 13F10 parasites expressing a PFA660–GFP fusion and labelled with Bodipy TR Ceramide to demarcate the PVM (other membranes are also labelled). TMP treatment as in **d, e**. All scale bars, 5 μ m. Images in **c–f** are representative of two independent experiments.

are secreted into the PV before entering the digestive vacuole where they are activated²⁵. Indeed, maturation of plasmepsin II was unaffected by TMP removal, indicating normal trafficking through the PV to the digestive vacuole (Extended Data Fig. 6). Additionally, HSP101 inactivation did not affect egress, a process mediated by secretion of the PFSUB1 protease into the PV²⁶. Because PV processes unrelated to export proceed normally when HSP101 function is interrupted, HSP101 seems to serve a specific role in export. Furthermore, parasites expressing a PV-targeted GFP^{DDD} fusion protein showed no defect in growth or export, indicating that the HSP101^{DDD} phenotype is not an indirect result of destabilizing the DDD in the PV compartment (Extended Data Fig. 7).

Finally, we examined PfGECO, a PEXEL-containing HSP40 that is specifically expressed in gametocytes²⁷. Again, export of PfGECO across the PVM was dramatically blocked in stage I gametocytes, showing an HSP101 requirement for protein export in sexual stages (Extended Data Fig. 4h).

Parasite nutrient acquisition is supported by establishment of a *Plasmodium* surface anion channel (PSAC or new permeability pathways) at the late ring stage. This channel modifies the sensitivity of the infected RBC to osmotic lysis upon solute uptake. Following TMP removal before RBC invasion, arrested ring-stage HSP101^{DDD} parasites failed to lyse in the presence of sorbitol (Fig. 3a). In contrast, sensitivity to osmotic lysis developed normally when TMP was removed at the end of the ring stage (Fig. 3b). CLAG3 proteins localize to the infected RBC periphery and are the only parasite proteins known to be involved in PSAC activation²⁸. Interestingly, we observed no change in CLAG3 localization to the host periphery when export was blocked, indicating that additional exported factors are required to activate PSAC (Fig. 3c). CLAG3 may be directly secreted into the RBC membrane during invasion similar to other rhoptry proteins²⁹, or could traffic from the PV by an HSP101-independent export pathway. Nutrient uptake via PSAC is an essential process and block of its formation at the late ring stage may be the lethal event that follows HSP101 inhibition.

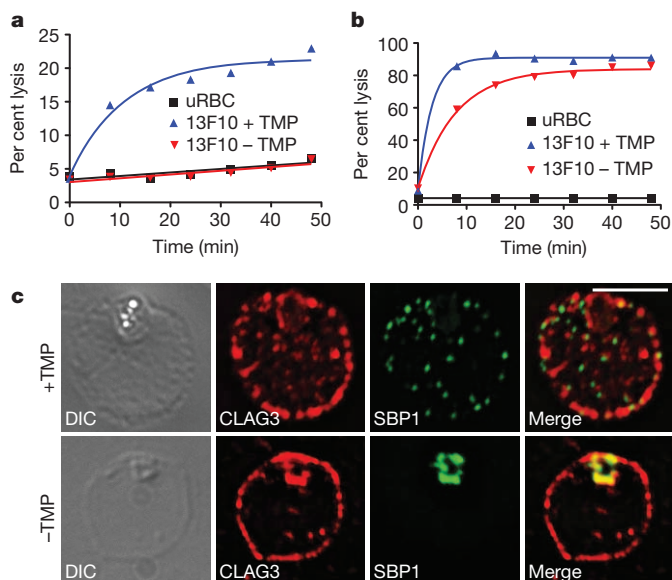


Figure 3 | HSP101 is required for activation of PSAC but not trafficking of CLAG3 to the RBC periphery. **a**, **b**, Osmotic lysis assay on 13F10 parasites. Sorbitol-sensitivity of arrested, late ring stage (–TMP) and control (+TMP) parasites at 25% parasitaemia (**a**) or later stages with or without TMP magnet-purified to >95% parasitaemia (**b**) is shown. Error bars represent s.d. of three technical replicates. Results are representative of two independent experiments. **c**, IFA of ring-stage 13F10 parasites showing CLAG3 localization to the RBC membrane with or without TMP. TMP was removed in the late schizont stage and parasites were allowed to develop 18 h before fixation with 90% acetone/10% methanol. Scale bar, 5 μ m. Images are representative of two independent experiments.

We reasoned that a block to protein export might arise from the disordered DDD interfering with interactions between HSP101^{DDD} and other PTEX components. To test this, HSP101^{DDD} was immunoprecipitated from parasite lysates. Remarkably, whereas localization of HSP101^{DDD} and EXP2 was unchanged with or without TMP (Extended Data Fig. 8a–c), their interaction decreased by more than 90% in the absence of TMP (Fig. 4a, f). A similar loss of HSP101–EXP2 interaction (70% decrease) was observed in reciprocal immunoprecipitation experiments (Fig. 4b, f). In contrast, association of HSP101^{DDD} with RESA increased by more than twofold (Fig. 4c, f and Extended Data Fig. 9a). To analyse PTEX150 interactions, we generated an HSP101^{DDD} strain expressing a second copy of PTEX150 with a Flag tag. The fusion protein localized to the PVM and distribution was unchanged with or without TMP (Extended Data Fig. 8d, e). Interaction between HSP101^{DDD} and PTEX150 was substantially decreased in the absence of TMP (84% decrease, Fig. 4d, f) whereas interactions between PTEX150 and EXP2 were not affected (Fig. 4e, f and Extended Data Fig. 9b). We conclude that destabilization of the DHFR-domain fold results in HSP101^{DDD} dissociation from the PTEX complex, producing a block in translocation of exported proteins out of the PV (Fig. 4g). Increased association with RESA under these conditions shows that interaction with exported substrates is not impaired. Interestingly, we also observed increased interaction between EXP2 and RESA in the absence of TMP (Fig. 4c, f and Extended Data Fig. 9a), suggesting that PTEX components can interact with substrates independent of HSP101 but that HSP101 is needed to drive export, probably through substrate unfolding and/or translocation (although we cannot exclude the possibility that association of RESA and HSP101^{DDD} in the absence of TMP results from ‘stickiness’ of the destabilized DDD).

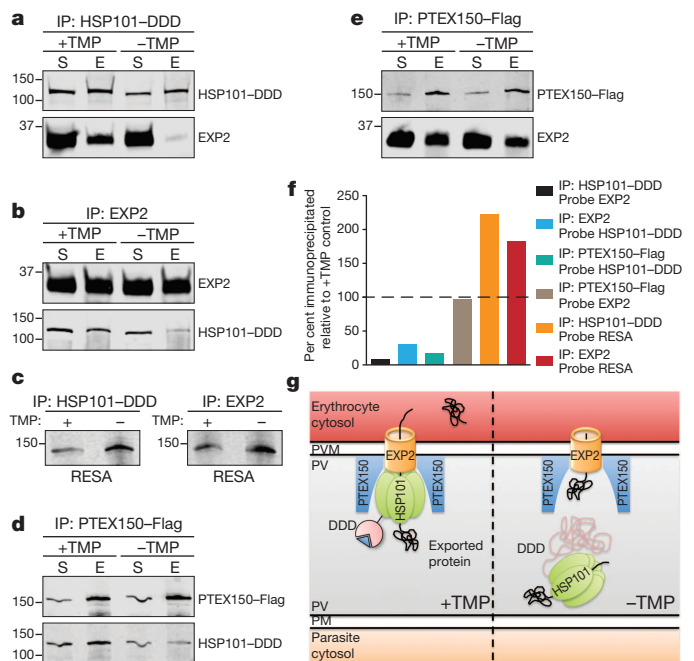


Figure 4 | Inactivated HSP101^{DDD} dissociates from the PTEX complex. **a**–**c**, Immunoprecipitation (IP) of PTEX components from lysates of 13F10 parasites. S, lysate supernatant input. E, elution. **d**, **e**, IP of PTEX components from lysates of 13F10 parasites expressing a PTEX150–Flag fusion protein. **f**, Quantification of IP data. Data are representative of three independent experiments. **g**, Model for mechanism of export block following inactivation of HSP101^{DDD}. HSP101 recognizes exported substrates and drives their translocation across the PVM in conjunction with other PTEX components by unfolding and/or directional threading. Destabilization of the DDD tag results in dissociation of HSP101 from PTEX, blocking translocation. Continued interaction of HSP101^{DDD} and EXP2 with exported substrates in these conditions suggests substrate molecules may be trapped at various steps in the translocation process.

These data demonstrate that HSP101 is required to mediate protein export in *P. falciparum*. As HSP101 is a major component of PTEX, our results provide strong functional evidence for a PTEX role in protein translocation (Fig. 4g), although we cannot formally rule out the possibility that these proteins perform some other critical activity upstream of actual translocation across the PVM. The essential role of HSP101 in parasite survival and export of key virulence determinants provides impetus for pursuing PTEX as a drug target.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 January; accepted 11 June 2014.

Published online 16 July; corrected online 30 July 2014 (see full-text HTML version for details).

- Boddey, J. A. & Cowman, A. F. *Plasmodium* nesting: remaking the erythrocyte from the inside out. *Annu. Rev. Microbiol.* **67**, 243–269 (2013).
- de Koning-Ward, T. F. *et al.* A newly discovered protein export machine in malaria parasites. *Nature* **459**, 945–949 (2009).
- Bullen, H. E. *et al.* Biosynthesis, localization, and macromolecular arrangement of the *Plasmodium falciparum* translocon of exported proteins (PTEX). *J. Biol. Chem.* **287**, 7871–7884 (2012).
- Riglar, D. T. *et al.* Spatial association with PTEX complexes defines regions for effector export into *Plasmodium falciparum*-infected erythrocytes. *Nature Commun.* **4**, 1415 (2013).
- Iwamoto, M., Björklund, T., Lundberg, C., Kirik, D. & Wandless, T. J. A general chemical method to regulate protein stability in the mammalian central nervous system. *Chem. Biol.* **17**, 981–988 (2010).
- Muralidharan, V., Oksman, A., Iwamoto, M., Wandless, T. J. & Goldberg, D. E. Asparagine repeat function in a *Plasmodium falciparum* protein assessed via a regulatable fluorescent affinity tag. *Proc. Natl Acad. Sci. USA* **108**, 4411–4416 (2011).
- Muralidharan, V., Oksman, A., Pal, P., Lindquist, S. & Goldberg, D. E. *Plasmodium falciparum* heat shock protein 110 stabilizes the asparagine repeat-rich parasite proteome during malarial fevers. *Nature Commun.* **3**, 1310 (2012).
- Silvestrini, F. *et al.* Protein export marks the early phase of gametocytogenesis of the human malaria parasite *Plasmodium falciparum*. *Mol. Cell. Proteomics* **9**, 1437–1448 (2010).
- Matthews, K. *et al.* The *Plasmodium* translocon of exported proteins (PTEX) component thioredoxin-2 is important for maintaining normal blood-stage growth. *Mol. Microbiol.* **89**, 1167–1186 (2013).
- Marti, M., Good, R. T., Rug, M., Knuefer, E. & Cowman, A. F. Targeting malaria virulence and remodeling proteins to the host erythrocyte. *Science* **306**, 1930–1933 (2004).
- Hiller, N. L. *et al.* A host-targeting signal in virulence proteins reveals a secretome in malarial infection. *Science* **306**, 1934–1937 (2004).
- Russo, I. *et al.* Plasmeprin V licenses *Plasmodium* proteins for export into the host erythrocyte. *Nature* **463**, 632–636 (2010).
- Boddey, J. A. *et al.* An aspartyl protease directs malaria effector proteins to the host cell. *Nature* **463**, 627–631 (2010).
- Rock, E. P. *et al.* Comparative analysis of the *Plasmodium falciparum* histidine-rich proteins HRP-I, HRP-II and HRP-III in malaria parasites of diverse origin. *Parasitology* **95**, 209–227 (1987).
- Spielmann, T. *et al.* A cluster of ring stage-specific genes linked to a locus implicated in cytoadherence in *Plasmodium falciparum* codes for PEXEL-negative and PEXEL-positive proteins exported into the host cell. *Mol. Biol. Cell* **17**, 3613–3624 (2006).
- Crabb, B. S. *et al.* Targeted gene disruption shows that knobs enable malaria-infected red cells to cytoadhere under physiological shear stress. *Cell* **89**, 287–296 (1997).
- Culvenor, J. G., Day, K. P. & Anders, R. F. *Plasmodium falciparum* ring-infected erythrocyte surface antigen is released from merozoite dense granules after erythrocyte invasion. *Infect. Immun.* **59**, 1183–1187 (1991).
- Boddey, J. A. *et al.* Role of plasmeprin V in export of diverse protein families from the *Plasmodium falciparum* exportome. *Traffic* **14**, 532–550 (2013).
- Grüring, C. *et al.* Uncovering common principles in protein export of malaria parasites. *Cell Host Microbe* **12**, 717–729 (2012).
- Mundwiler-Pachlatko, E. & Beck, H. P. Maurer's clefts, the enigma of *Plasmodium falciparum*. *Proc. Natl Acad. Sci. USA* **110**, 19987–19994 (2013).
- Blisnick, T. *et al.* Pfsbp1, a Maurer's cleft *Plasmodium falciparum* protein, is associated with the erythrocyte skeleton. *Mol. Biochem. Parasitol.* **111**, 107–121 (2000).
- Heiber, A. *et al.* Identification of new PNEPs indicates a substantial non-PEXEL exportome and underpins common features in *Plasmodium falciparum* protein export. *PLoS Pathog.* **9**, e1003546 (2013).
- Miller, L. H., Baruch, D. I., Marsh, K. & Dombro, O. K. The pathogenic basis of malaria. *Nature* **415**, 673–679 (2002).
- Külzer, S. *et al.* Parasite-encoded Hsp40 proteins define novel mobile structures in the cytosol of the *P. falciparum*-infected erythrocyte. *Cell. Microbiol.* **12**, 1398–1420 (2010).
- Klemba, M., Beatty, W., Gluzman, I. & Goldberg, D. E. Trafficking of plasmeprin II to the food vacuole of the malaria parasite *Plasmodium falciparum*. *J. Cell Biol.* **164**, 47–56 (2004).
- Yeoh, S. *et al.* Subcellular discharge of a serine protease mediates release of invasive malaria parasites from host erythrocytes. *Cell* **131**, 1072–1083 (2007).
- Morahan, B. J. *et al.* Functional analysis of the exported type IV HSP40 protein PfGECO in *Plasmodium falciparum* gametocytes. *Eukaryot. Cell* **10**, 1492–1503 (2011).
- Nguitragool, W. *et al.* Malaria parasite *clag3* genes determine channel-mediated nutrient uptake by infected red blood cells. *Cell* **145**, 665–677 (2011).
- Besteiro, S., Michelin, A., Poncet, J., Dubremetz, J. F. & Lebrun, M. Export of a *Toxoplasma gondii* rhoptry neck protein complex at the host cell membrane to form the moving junction during invasion. *PLoS Pathog.* **5**, e1000309 (2009).

Acknowledgements This work was supported by National Institutes of Health grants AI047798 to D.E.G., T32-AI007172 to J.R.B. and AI099156 to V.M. We thank J. McBride, D. Cavanagh and EMRR for anti-EXP2 antibody, J. Adams and ATCC (MR4) for anti-BIP antibody, D. Taylor for anti-HRP2 antibody, R. Anders for anti-RESA antibody, C. Braun-Breton for anti-SBP1 antibody, K. Williamson for anti-PfGECO and anti-Pfs16 antibodies, T. Spielmann for anti-REX2, anti-REX3 and anti-MSRP6 antibodies, L. Tilley for anti-REX1 and anti-PfEMP1 antibodies, S. Desai for anti-CLAG3 antibody, A. Cowman for anti-KAHRP antibody, J. Przyborski and K. Lingelbach for anti-SERP antibody, W. Beatty for assistance with electron microscopy, B. Vaupel and T. Butler for technical assistance and P. Sigala and N. Spillman for suggestions.

Author Contributions J.R.B., V.M. and D.E.G. conceived and designed experiments. J.R.B. performed the majority of the experiments and V.M. performed some experiments. V.M. and A.O. generated the HSP101^{DD} strains. J.R.B. and A.O. performed the gametocyte analysis. J.R.B. and D.E.G. analysed the data and wrote the manuscript. All authors discussed and edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.E.G. (dgoldberg@wustl.edu).

METHODS

Cell fractionation. TMP was removed from synchronous 13F10 parasites at the late schizont stage and development was allowed to proceed with or without 10 μ M TMP for 18 h to the ring stage. Ring-stage parasites were enriched to >95% parasitaemia by an initial tetanolyisin treatment of 2.5 haemolytic units (HU) for 20 min as described³⁰. Samples were then fractionated by sequential treatment with tetanolyisin (1 HU for 20 min) and saponin (0.035% for 5 min) as described^{19,30}.

Parasite strains, culture and genetic modification. Parasite culture and flow cytometry analysis was performed as described⁷. Cloning was performed using the In-Fusion system (Clontech) unless otherwise noted. For generation of HSP101^{DDD} parasites, the *DDD* cassette was PCR-amplified from plasmid pGDB⁶ with primers 5'-GAGCAGCTGACCTAGGTACCCATACGATGTTCCAGATTACGCTTACC CATAAGTGTCCAGATTACGCTTACCCATACGATGTTCCAGATTACGCTTATACAGTCTGATTGCGCGGTAGCGGTATACGATGTTACGCG-3' and 5'-TAACTCGACGCGCGGTCATCGCGCTCCAGAACTCTCAAAGCAATAGCTGTGAGAG-3' to incorporate a 5' 3 \times HA tag. The amplicon was inserted into plasmid pPM2GT²⁵ between restriction sites AvrII and EagI. The *BSD* cassette was then PCR-amplified from pGDB with primers 5'-ATGCCTGCAGGTGCGATTGCATGCTTAGCTAATTCGCTG-3' and 5'-GAGAACTCTAGAGGTACCGAGCTCGATCTGCCGCTCTCC-3' and inserted between restriction sites SalI and BglII. Finally, a portion of the 3' genomic locus of *HSP101* up to but not including the stop codon was PCR-amplified from *P. falciparum* genomic DNA using primers 5'-ACGATTTTCTCGAGCGAAACCTTTTATGTTATTAATATAACAGATAAAGCTTTAGTACGACAGC-3' and 5'-CTGCACCTGGCCTAGGGTCTTAGATAAGTTTATAACCAAGTTTCTAGCTTTACTATTATAATCAACAAATACATCC-3' and inserted between restriction sites XhoI and AvrII. The resulting plasmid pHSP101-HDB was transfected into the 3D7-derived parental strain PMIKO³¹, which contains an hDHFR expression cassette conferring resistance to TMP. Selection, cycling and cloning were performed as described⁷. TMP was maintained in the medium throughout. Two clones from independent transfections were isolated and designated 13F10 and 14G11. Integration was detected by PCR using the primers 5'-CCTCCTTCAGTAGATATGACCG-3' and 5'-CTAACGCCGCAATCAGACTG-3' and confirmed by Southern blot using a probe to the 3' end of the *HSP101* gene, generated with the primers 5'-CGAAACCTTTTATGTTATTAATATAACAG-3' and 5'-GGTCTTAGATAAGTTTATAACCAAG-3'.

For gene expression, the episomal expression plasmid pTyEOE⁷ containing a yeast dihydroorotate dehydrogenase selection cassette was modified for transposase-mediated genomic integration by inserting a *piggyBac* element containing inverted terminal repeats amplified from plasmid pXL-BACII-DHFR³² with primers 5'-GG AATTTCTTATAAGATCTTAATACGACTCACTATAGGGCGAATTGGG-3' and 5'-ATAATGGTTTCTTAGACGTCGATAAAAGTTTGTACTTTATAGAAGAAATTTTGTAG-3' between the restriction sites BglII and AatII, resulting in the plasmid pTyEOE. For expression of PFA660-GFP, the PFA660 coding sequence was PCR-amplified from parasite complementary DNA with primers 5'-ACGATTTTCTCGAGATGGCAACCTTAAGGAAAAGC-3' and 5'-CTGCACCTGGCCTAGGATACTCTCTTAAATATCTCTTTATC-3' and inserted into plasmid pTyEOE between restriction sites XhoI and AvrII, placing the gene under the control of the HSP86 promoter and generating a C-terminal GFP fusion. For expression of the PV-targeted GFP-DDD fusion protein, sequence encoding GFP fused to DDD with a C-terminal HA tag was amplified from plasmid pGDB with the forward primer 5'-ACCCCGGGATCTCGAGATGACAAGAAGATATTTAAAGTATTATTTTGTGTTACTTTATTGTTTGTTCAGTTATTAATAATGTAATTGTGTGCTCCTAGGGCAGCAAGTAAAGGAGAAGAACTTTTCACTGGAG-3' (encoding the HSP101 signal peptide, amino acids 1–27) and the reverse primer 5'-TAACTCGACGCGCGCTCAAGCGTAATCTGGAACATCGTATGGG-3'. This amplicon was inserted in plasmid pTyEOE between the restriction sites XhoI and EagI. For expression of the PTEX150-Flag fusion protein, sequence encoding a 6 \times Flag epitope was generated with the primer 5'-CTGTAGACCTAGGACTACAAGACGACGACGACAAGGATTATAAAGATGATGATGATAAAGATTATAAAGATGATGATGATAAAGATTATAAAGATGATGATGATAAATAACGGCCGCTCGAGTT-3' and its reverse complement. The primers were annealed and inserted between restriction sites AvrII/EagI in plasmid pTyEOE, resulting in plasmid pTyEOE-6xFlag. The *PTEX150* gene (which contains no introns) was then PCR-amplified from parasite genomic DNA with its endogenous 5' UTR up to but not including the stop codon using primers 5'-GATCGAGACGTCCTCTTGTGGTCAAAATAAGTAAATTTTATAAATTC-3' and 5'-GTCACCTAGGATTGTCGTCCTCTTCTTCGTCC-3' and inserted between restriction sites AatII/AvrII by standard ligation. pTyEOE-derived plasmids were co-transfected with plasmid pHTH for transient expression of the *piggyBac* transposase to mediate genomic integration³².

Growth assays. For asynchronous growth assays, parasites were subcultured 1:2 each day to avoid confounding effects from high parasite density in making growth rate comparisons between different samples. Parasitaemia at each time point was

back calculated based on the subculturing schedule. Data were fit to exponential growth equations using Prism (GraphPad Software, Inc.). In synchronous growth assays, daily media changes were performed without subculture. Parasitaemia was determined by flow cytometry.

Gametocyte induction. Synchronous 13F10 parasites at 5–7% parasitaemia were stressed by increasing haematocrit to 4% at the mid-trophozoite stage for ~12 h. Haematocrit was returned to 2% at the schizont stage and TMP washout was performed at late schizont stage shortly before rupture and at 24 h intervals thereafter. Asexual parasites were killed by treatment with 50 mM N-acetylglucosamine in the following cycle.

Antibodies. The following antibodies were used for IFA and western blot (WB): rabbit polyclonal anti-HA (Life Technologies) (IFA: 1:100); rabbit polyclonal anti-HA (Sigma) (WB 1:1,000); rat anti-HA mAb clone 3F10 (Roche) (IFA: 1:100, WB: 1:1,000); mouse anti-Flag mAb clone M2 (Sigma) (IFA: 1:100, WB 1:500); mouse anti-EXP2 mAb clone 7.7³³ (IFA: 1:1,000, WB: 1:1,000); mouse anti-HRP2 mAb 2G12¹⁴ (IFA 1:1,000, WB: 1:500); mouse anti-RESA mAb clone 28/2³⁴ (IFA: 1:1,000, WB: 1:1,000); rabbit polyclonal anti-BiP³⁵ (WB: 1:1,000); rabbit polyclonal anti-REX1³⁶ (IFA: 1:1,000); mouse polyclonal anti-REX2¹⁵ (IFA: 1:500); mouse polyclonal anti-REX3¹⁵ (IFA: 1:500); rabbit polyclonal anti-SBP1²¹ (IFA: 1:500); mouse polyclonal anti-MSRP6²² (IFA: 1:500); mouse polyclonal anti-PfGECO²⁷ (IFA: 1:300); rabbit polyclonal anti-Pf16³⁷ (IFA: 1:500); rabbit polyclonal anti-PfEMP1 ATS³⁸ (IFA: 1:200); mouse polyclonal anti-CLAG3³⁸ (IFA: 1:300); rabbit polyclonal anti-PM2 antibody 737³⁹ (WB 1:2,000); rabbit polyclonal anti-KAHRP¹⁶ (IFA: 1:500); rabbit polyclonal anti-SERP⁴⁰ (IFA: 1:500, WB: 1:1,000).

Light microscopy and image processing. For indirect IFA detecting HRP2, REX3, PfGECO and Pf16, cells were allowed to settle on coverslips coated with concanavalin A and fixed with a mixture of 4% paraformaldehyde and 0.0075% glutaraldehyde in PBS. For indirect IFA detecting HA and Flag epitopes, EXP2, RESA, REX1, REX2, SBP1, MSRP6, PfEMP1, and KAHRP, thin smears were air dried and fixed with room temperature acetone. For indirect IFA detecting CLAG3 and SBP1, thin smears were air dried and fixed with ice-cold 90% acetone/10% methanol. Primary antibodies were detected by Alexa Fluor 488 or 594 secondary IgG antibodies (Life Technologies) used at 1:2,000. Image stacks were collected at z-increments of 0.2 μ m with an ORCA-ER CCD camera (Hamamatsu) and AxioVision software on an Axio Imager.M1 microscope (Zeiss) using a \times 100 oil immersion objective. Deconvolved images were generated using manufacturer specified point-spread functions and displayed as maximum intensity projections. Adjustments to brightness and contrast were made for display purposes. Live imaging was performed as previously described⁴¹. All IFA and live imaging experiments were repeated at least twice.

Immunoelectron microscopy. For immunolocalization at the ultrastructural level, infected RBCs were fixed in 4% paraformaldehyde/0.05% glutaraldehyde (Polysciences Inc.) in 100 mM PIPES/0.5 mM MgCl₂, pH 7.2 for 1 h at 4 °C. Samples were then embedded in 10% gelatin and infiltrated overnight with 2.3 M sucrose/20% polyvinyl pyrrolidone in PIPES/MgCl₂ at 4 °C. Samples were trimmed, frozen in liquid nitrogen, and sectioned with a Leica Ultracut UCT cryo-ultramicrotome (Leica Microsystems Inc.). 50-nm sections were blocked with 5% fetal bovine serum/5% normal goat serum for 30 min and subsequently incubated with primary antibodies followed by secondary anti-rabbit conjugated to 18 nm colloidal gold (Jackson ImmunoResearch Laboratories, Inc.). Sections were washed in PIPES buffer followed by a water rinse, and stained with 0.3% uranyl acetate/2% methyl cellulose. Samples were viewed with a JEOL 1200EX transmission electron microscope (JEOL USA Inc.) equipped with an AMT 8 megapixel digital camera (Advanced Microscopy Techniques). All labelling experiments were conducted in parallel with controls omitting the primary antibody. These controls were consistently negative at the concentration of colloidal gold conjugated secondary antibodies used in these studies.

Assessment of export in recently invaded parasites. To monitor RESA export immediately after invasion, synchronous late schizonts were purified over Percoll and then mixed with fresh RBCs in pre-warmed media. Cultures were shaken at 37 °C for one hour. After lysis of unruptured schizonts by sorbitol treatment, samples were fixed for IFA and stained with RESA antibodies as described above.

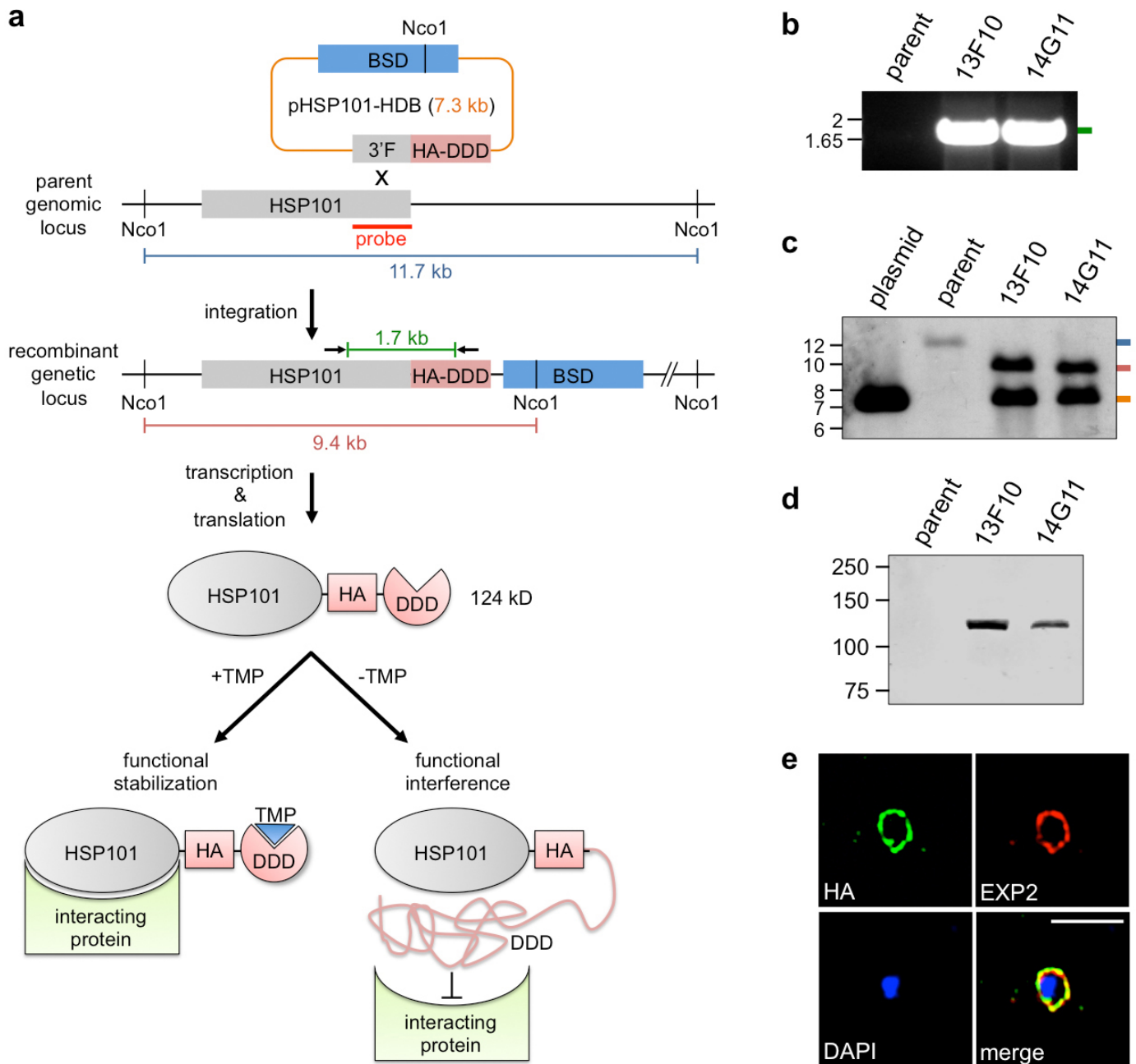
Osmotic lysis assays. Osmotic lysis assays were performed as described by incubating cells in 290 mM D-sorbitol and measuring hemoglobin release by absorbance at 405 nm^{42,43}. Samples were prepared as follows: TMP was removed or not from synchronous, late schizont-stage 13F10 parasites and development was allowed to proceed for 24 h, resulting in 25% parasitaemia cultures of early trophozoites (+TMP) or arrested late ring-forms (-TMP). Alternatively, TMP was removed or not from synchronous 13F10 parasites at the late ring-stage and development was allowed to proceed 24 h before magnet purification to >95% parasitaemia. In either case, the final concentration of cells in each sample was 1 \times 10⁸/ml. At the end of the assay, samples were lysed with saponin and data were normalized to these values (100% lysis).

Immunoprecipitation. For immunoprecipitation, TMP was removed or not from synchronous parasites at the early trophozoite-stage and parasites were allowed to

develop for 24 h to late schizonts and early rings. Parasites were then harvested with 0.035% saponin in PBS and lysed in 0.5% triton X-100 in PBS. Following sonication, lysates were cleared by centrifugation at 16,000g for 5 min and supernatants were nutated for 3 h at 4 °C with indicated antibodies and Dynabeads coupled to protein G or protein A (or Dynabeads alone as a negative control when necessary) (Life Technologies). Immune complexes were purified on magnets with extensive washing with 0.5% triton X-100 in PBS and then lysed in sample buffer.

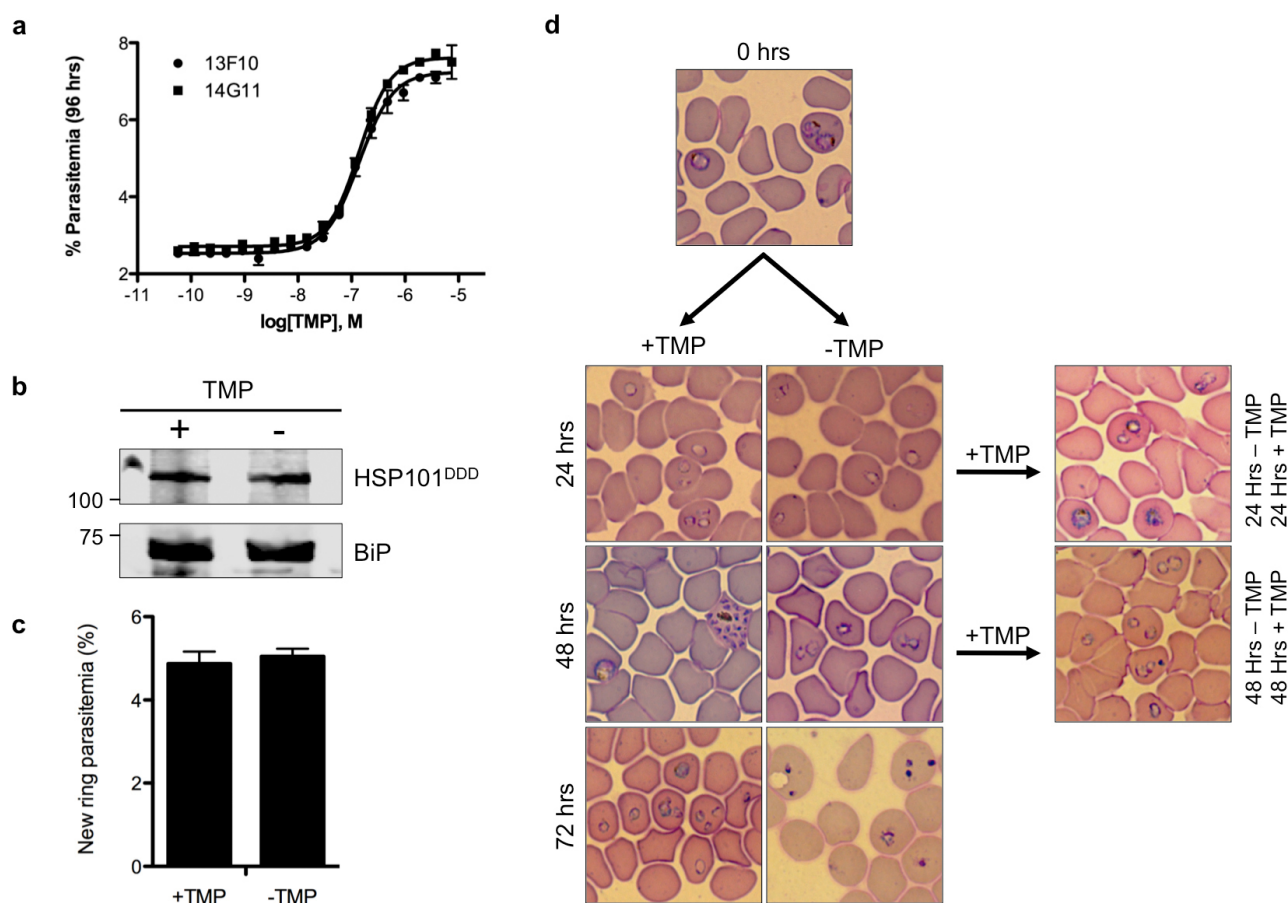
Western blots. Western blot analysis was performed using an Odyssey infrared imaging system (LI-COR Biosciences). Primary antibodies were detected by IRDye 680 and 800-conjugated secondary antibodies (LI-COR Biosciences) used at 1:15,000. Quantitative measurements were made with Image Studio software (LI-COR Biosciences).

30. Jackson, K. E. *et al.* Selective permeabilization of the host cell membrane of *Plasmodium falciparum*-infected red blood cells with streptolysin O and equinatoxin II. *Biochem. J.* **403**, 167–175 (2007).
31. Liu, J., Gluzman, I. Y., Drew, M. E. & Goldberg, D. E. The role of *Plasmodium falciparum* food vacuole plasmepsins. *J. Biol. Chem.* **280**, 1432–1437 (2005).
32. Balu, B., Shoue, D. A., Fraser, M. J., Jr & Adams, J. H. High-efficiency transformation of *Plasmodium falciparum* by the lepidopteran transposable element *piggyBac*. *Proc. Natl Acad. Sci. USA* **102**, 16391–16396 (2005).
33. Hall, R. *et al.* Antigens of the erythrocytes stages of the human malaria parasite *Plasmodium falciparum* detected by monoclonal antibodies. *Mol. Biochem. Parasitol.* **7**, 247–265 (1983).
34. Anders, R. F. *et al.* in *Molecular Strategies of Parasitic Invasion* (eds Goodman, H., Agabian, N. & Noguiera, N.) 333–342 (Alan R. Liss, 1987).
35. Kumar, N., Koski, G., Harada, M., Aikawa, M. & Zheng, H. Induction and localization of *Plasmodium falciparum* stress proteins related to the heat shock protein 70 family. *Mol. Biochem. Parasitol.* **48**, 47–58 (1991).
36. Hawthorne, P. L. *et al.* A novel *Plasmodium falciparum* ring stage protein, REX, is located in Maurer's clefts. *Mol. Biochem. Parasitol.* **136**, 181–189 (2004).
37. Eksi, S. *et al.* Identification of a subtelomeric gene family expressed during the asexual-sexual stage transition in *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **143**, 90–99 (2005).
38. Frankland, S. *et al.* Serum lipoproteins promote efficient presentation of the malaria virulence protein PfEMP1 at the erythrocyte surface. *Eukaryot. Cell* **6**, 1584–1594 (2007).
39. Francis, S. E., Banerjee, R. & Goldberg, D. E. Biosynthesis and maturation of the malaria aspartic hemoglobins plasmepsins I and II. *J. Biol. Chem.* **272**, 14961–14968 (1997).
40. Ragge, K. *et al.* *In vitro* biosynthesis and membrane translocation of the serine rich protein of *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **42**, 93–100 (1990).
41. Grüning, C. & Spielmann, T. Imaging of live malaria blood stage parasites. *Methods Enzymol.* **506**, 81–92 (2012).
42. Ginsburg, H., Kutner, S., Krugliak, M. & Cabantchik, Z. I. Characterization of permeation pathways appearing in the host membrane of *Plasmodium falciparum* infected red blood cells. *Mol. Biochem. Parasitol.* **14**, 313–322 (1985).
43. Kirk, K., Horner, H. A., Elford, B. C., Ellory, J. C. & Newbold, C. I. Transport of diverse substrates into malaria-infected erythrocytes via a pathway showing functional characteristics of a chloride channel. *J. Biol. Chem.* **269**, 3339–3347 (1994).
44. Babbitt, S. E. *et al.* *Plasmodium falciparum* responds to amino acid starvation by entering into a hibernatory state. *Proc. Natl Acad. Sci. USA* **109**, E3278–E3287 (2012).



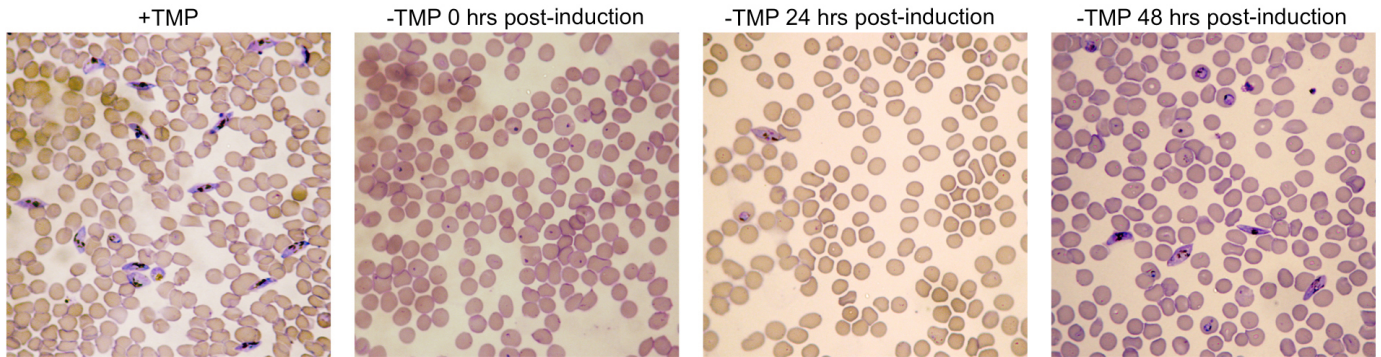
Extended Data Figure 1 | Generation of HSP101^{DDD} strains. **a**, Schematic of strategy to generate HSP101^{DDD} parasites. 3'F, 3' flank for homologous recombination; BSD, blasticidin S deaminase; TMP, trimethoprim; HA, haemagglutinin tag; DDD, DHFR destabilization domain. **b**, Diagnostic PCR showing integration of the DDD fusion in the two independent clones 13F10 and 14G11. Primers shown as black arrows in **a**. Image is representative of two independent experiments. **c**, Southern blot showing integration of plasmid pHSP101-HDB occurred at the intended genomic locus. Expected NcoI

digestion products and sizes are indicated in blue, red and orange in **a**. The 7.3-kb band in 13F10 and 14G11 indicates the presence of concatemers commonly observed in *P. falciparum*. Image is representative of one experiment. **d**, Western blot with anti-HA antibodies detects a 124 kDa band in clones 13F10 and 14G11. Image is representative of two independent experiments. **e**, IFA of acetone-fixed 13F10 parasites showing colocalization of HSP101^{DDD} and EXP2 at the PVM. Scale bar, 5 μ m. Images are representative of two independent experiments.



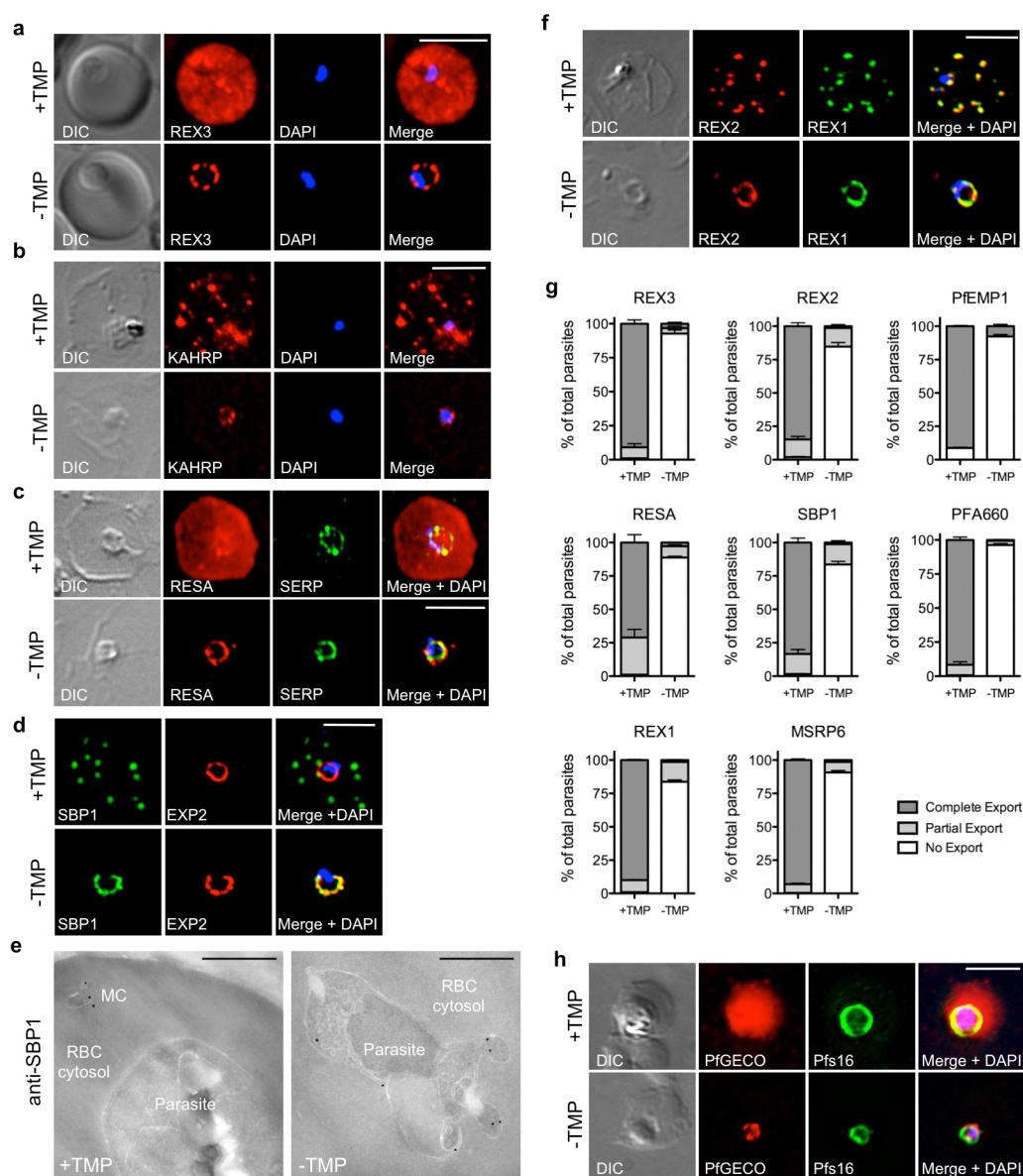
Extended Data Figure 2 | Characterization of HSP101^{DDD} strains. **a**, TMP dose-response for 13F10 and 14G11 parasites grown for 96 h was measured. Error bars represent s.d. of three technical replicates. **b**, Western blot on lysates from asynchronous 13F10 parasites grown with or without TMP for 24 h. BiP serves as a loading control. Blotting with anti-HA antibody shows no decrease in HSP101^{DDD} protein levels relative to the BiP loading control. Images are representative of two independent experiments. **c**, Quantification of new ring parasitaemia with or without TMP. In a parallel experiment to that shown in Fig. 1d (but beginning with lower parasitaemia cultures), new rings were counted in Giemsa-stained smears of synchronous 13F10 parasites where TMP

was removed or not during the proceeding trophozoite stage. No significant difference in the resulting new ring parasitaemia was observed with or without TMP, indicating no difference in re-invasion efficiency-TMP. Error bars represent s.d. of three technical replicates. Data are representative of two independent experiments. **d**, Giemsa-stained smears of synchronous 13F10 parasites grown with or without TMP from Fig. 1d. Reintroduction of TMP after 24 or 48 h restored progression through the intraerythrocytic cycle. After 72 h without TMP, most parasites appeared as dead, pyknotic forms. Images are representative of three independent experiments. Original magnification $\times 1,000$.



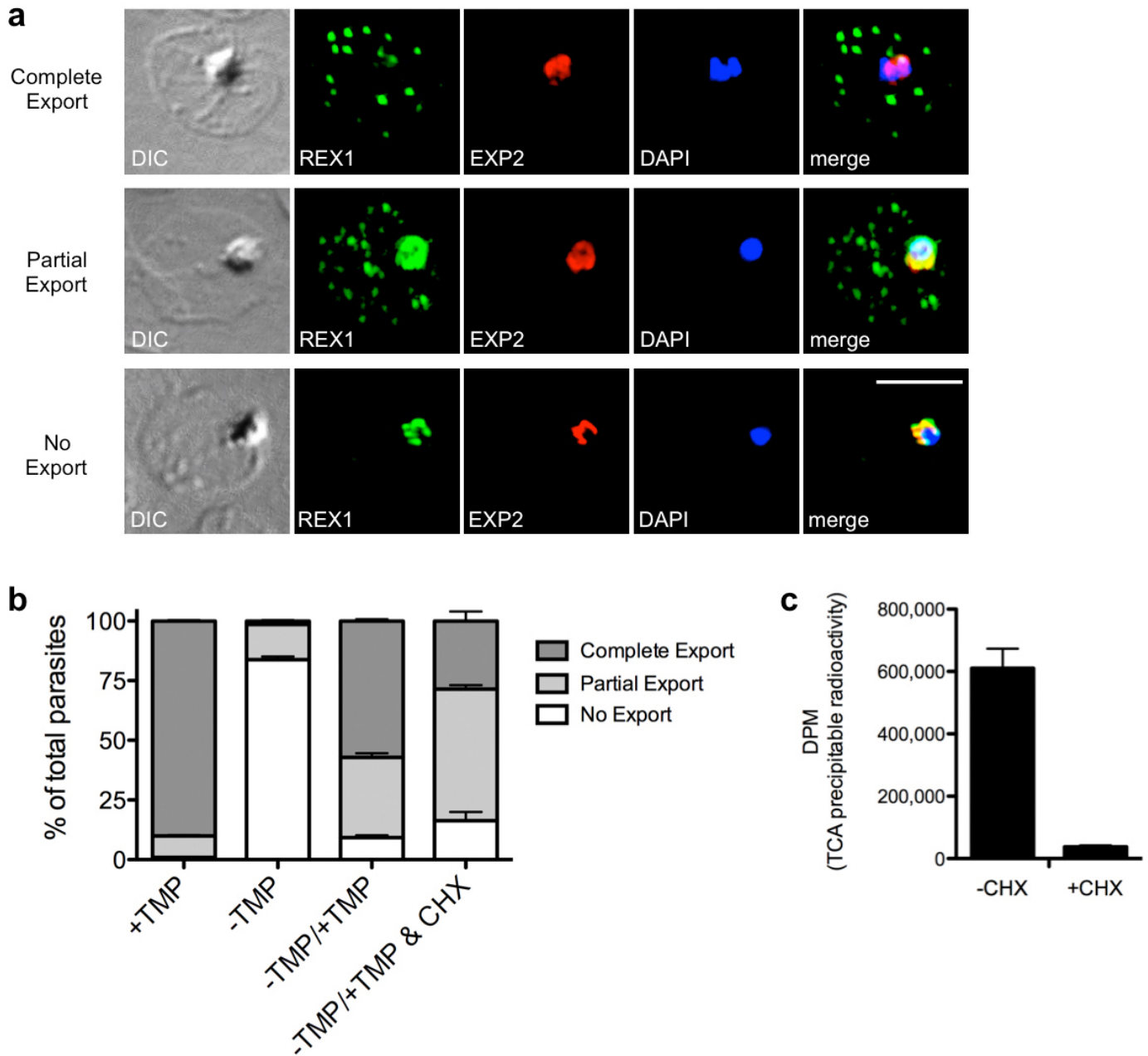
Extended Data Figure 3 | HSP101 function is critical for early stage gametocyte development. Giemsa-stained smears from day nine post gametocyte induction as quantified in Fig. 1e. TMP was removed from parallel samples at 24-h intervals beginning just before reinvasion (0 h) following

gametocytogenesis induction. Data are representative of 4 independent experiments. While control (+TMP) gametocytaemia varied between experiments, a similar effect on gametocyte formation was observed in each experiment following TMP removal. Original magnification $\times 1,000$.



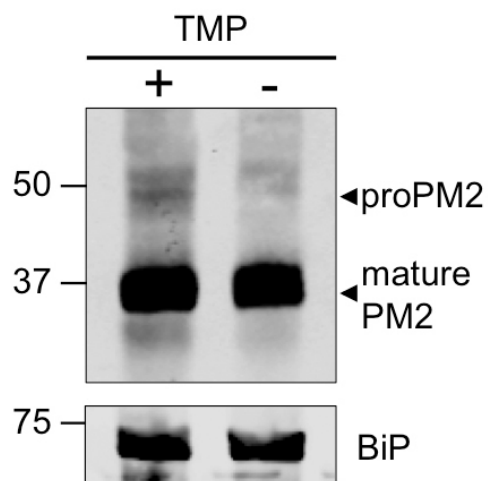
Extended Data Figure 4 | Localization and quantification of exported proteins following HSP101^{DDD} inactivation. **a–f**, IFA and immunoelectron microscopy of ring-stage 13F10 parasites with or without TMP. TMP was removed in late schizont stage and parasites were allowed to reinvasion and grow 18–24 h before fixation with paraformaldehyde (**a**) or acetone (**b–d, f**). Immunoelectron microscopy fixation (**e**) is detailed in methods. A similar export block with accumulation of the exported protein at the parasite periphery was observed in each case when TMP was removed. The soluble PEXEL-containing protein REX3 (**a**) is normally exported into the host RBC cytosol. The PEXEL-containing KAHRP protein (**b**) is normally exported through Maurer's clefts to knob structures at the cytoplasmic face of the infected RBC membrane. The PEXEL-containing protein RESA (**c**) is normally exported to the RBC periphery. SERP is a marker for the PV. SBP1 (**d**) is an integral membrane PNEP normally exported to the Maurer's clefts. In the absence of TMP, blocked SBP1 colocalizes with EXP2. **e**, Immunoelectron microscopy showing localization of SBP1 in ring-stage parasites with or

without TMP. MC, Maurer's cleft. Scale bars, 500 nm. Images are representative of one experiment. REX2 (**f**) is an integral membrane PNEP normally exported to the Maurer's clefts. **g**, Quantification of export block by IFA for exported proteins shown here and in Fig. 2c–f. Export was scored as complete (all signal in the host cell), partial (signal within the host cell but also within the PV) or no export (signal only seen within the PV and not in the host cell). Example images of each scoring scenario are given for REX1 in Extended Data Fig. 5a. In the case of PfEMP1, cells were scored as having PfEMP1 signal at the RBC periphery (complete export) or not (no export) due to the fact that some PfEMP1 signal is always seen within the PV under normal export conditions (see Fig. 2e, +TMP). Error bars represent s.d. of three technical replicates. Data are representative of at least two independent experiments. **h**, IFA of PfGECO in paraformaldehyde-fixed, stage I gametocytes 36 h post invasion with or without TMP. Pfs16 is a gametocyte-specific PVM marker. All IFA scale bars, 5 μ m. All IFA images (**a–d, f, h**) are representative of two independent experiments.

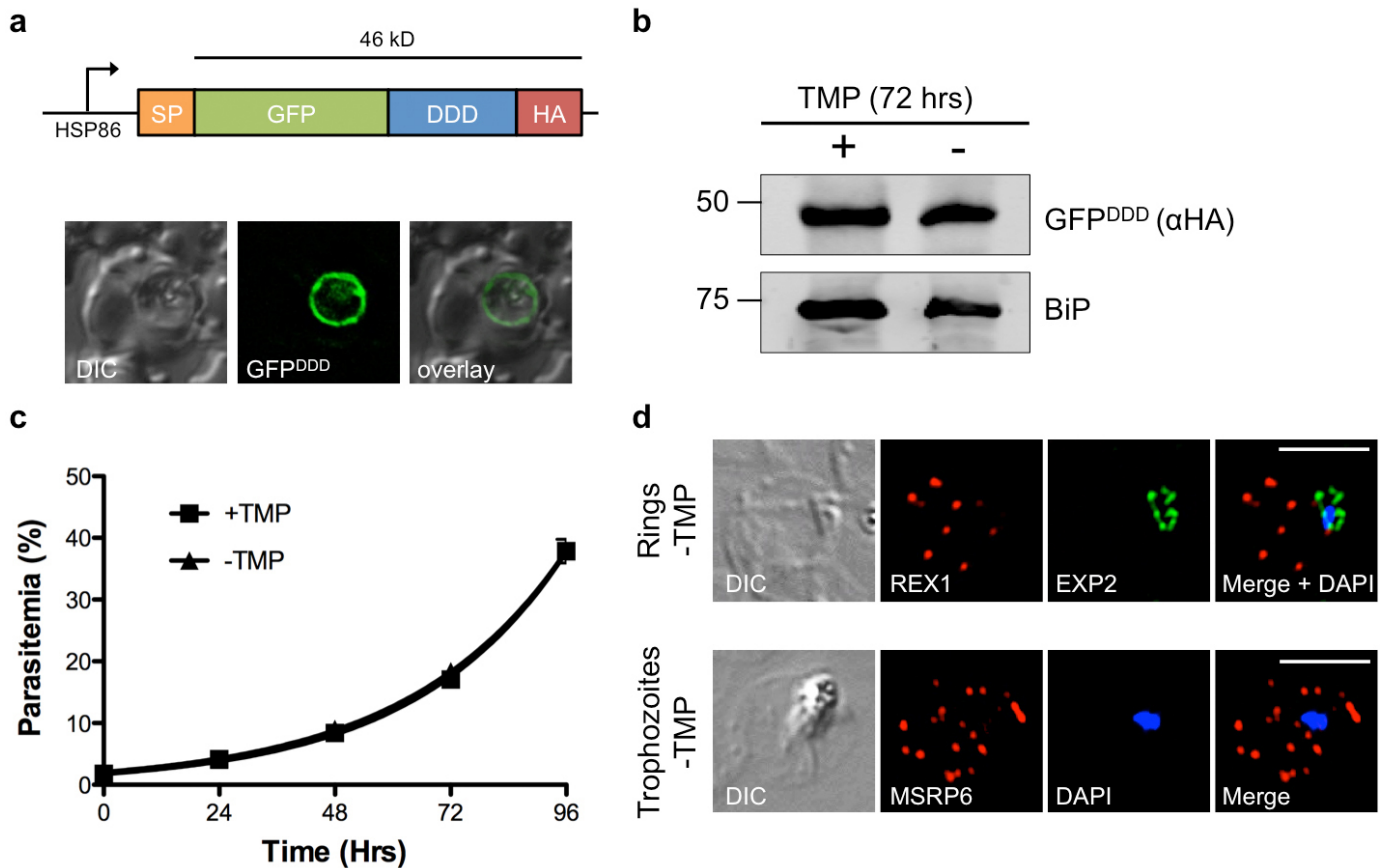


Extended Data Figure 5 | Reactivation of export does not require new protein synthesis. **a, b,** TMP was removed in late schizonts and parasites were allowed to reinvade and develop for 18 h before TMP add back with or without $10 \mu\text{g ml}^{-1}$ cycloheximide (CHX). Parasites were acetone-fixed 24 h later and processed for IFA. Export was scored as complete (no REX1 retained within the EXP2-labelled PVM), partial (REX1 in the host cell and retained within the PVM) or no export (no REX1 signal beyond the PVM). Error bars

represent s.d. of three technical replicates. Data are representative of two independent experiments. Scale bar, $5 \mu\text{m}$. **c,** Metabolic labelling with [^{35}S]methionine/cysteine, performed as previously described⁴⁴, confirms that CHX treatment conditions inhibit new protein synthesis. Parasite proteins were TCA-precipitated and incorporated radioactivity was determined through scintillation counting. Error bars represent s.d. of three technical replicates.

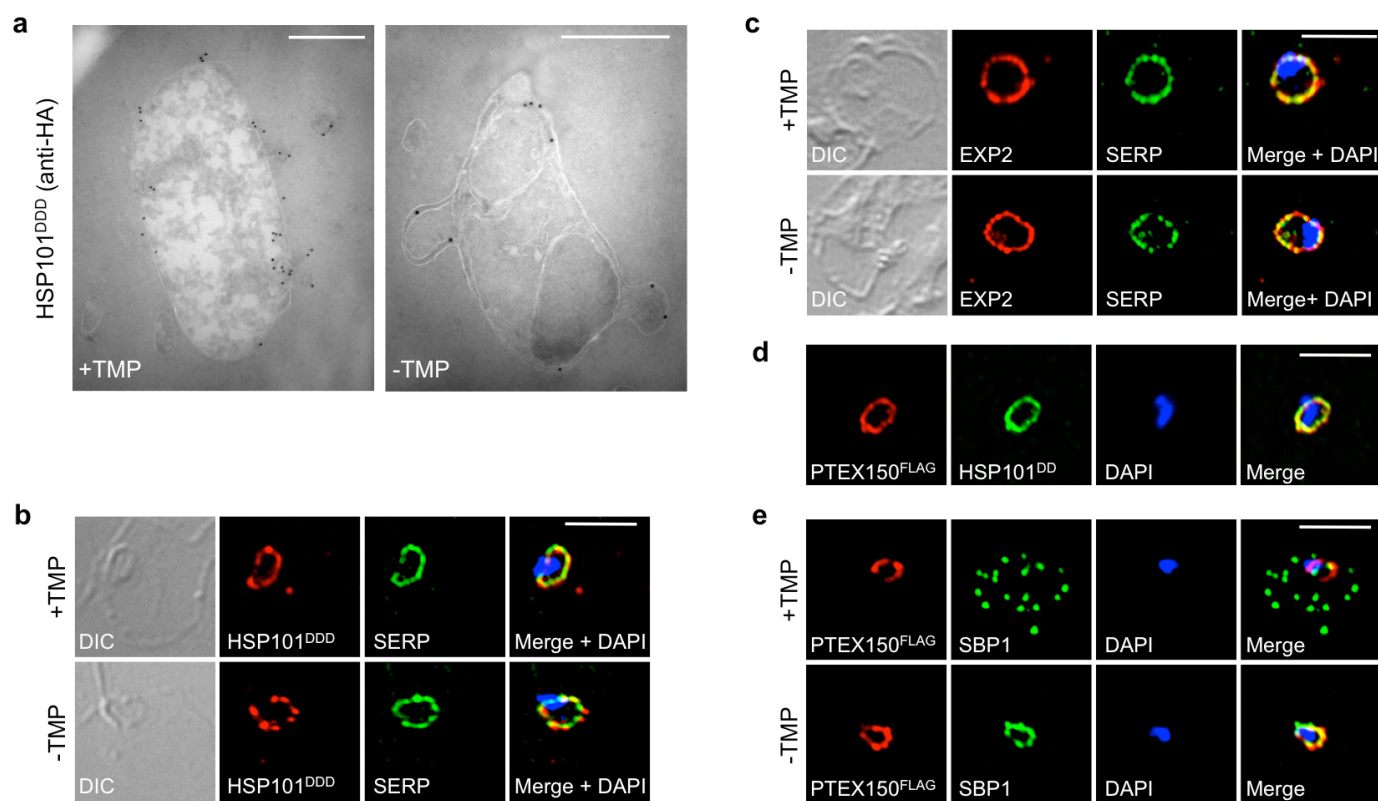


Extended Data Figure 6 | Maturation of plasmepsin 2 is not affected by inactivation of HSP101. Western blot showing normal maturation of plasmepsin 2 (PM2) in asynchronous parasites after 24 h –TMP. BiP serves as a loading control. Maturation requires proPM2 trafficking through the PV before internalization to the digestive vacuole where maturation occurs²⁵. Images are representative of two independent experiments.



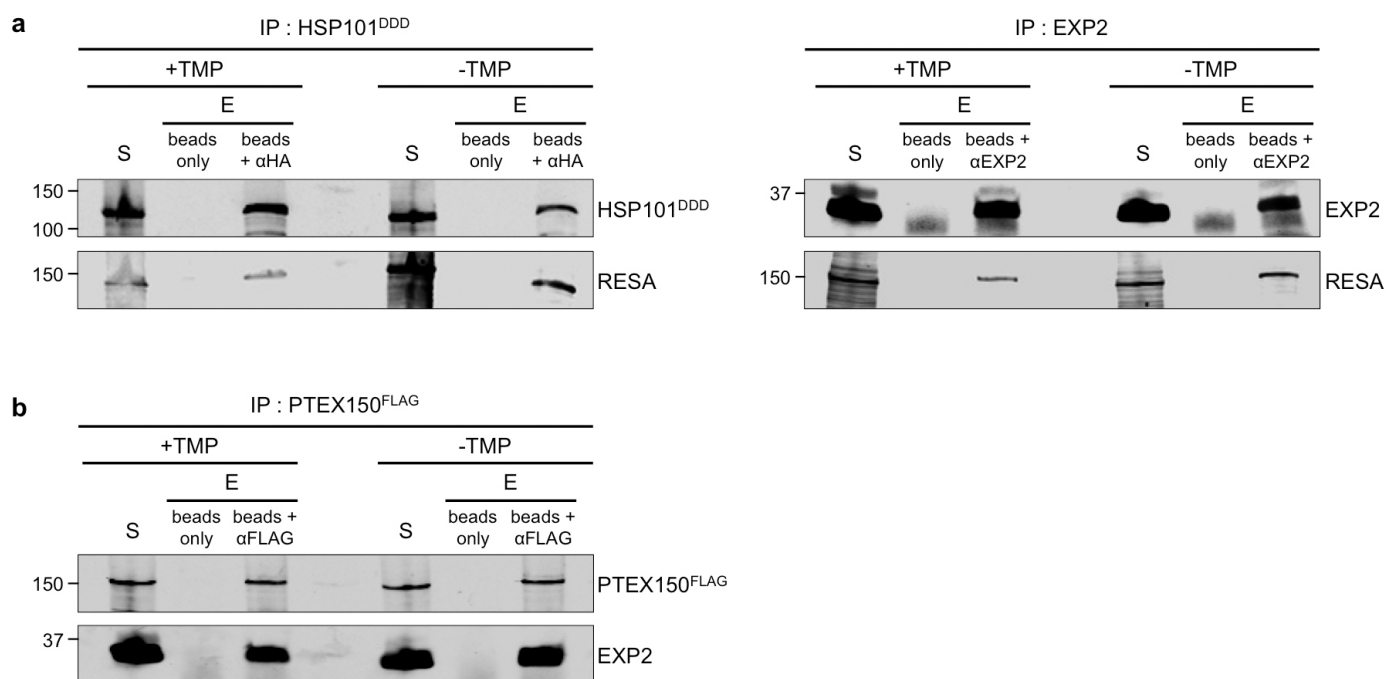
Extended Data Figure 7 | DDD targeted to the PV independent of HSP101 does not interfere with parasite growth or export. **a**, Schematic of the PV-targeted GFP-DDD fusion protein consisting of a signal peptide appended to a GFP-DDD fusion with a C-terminal HA epitope tag and expressed under the control of the HSP86 promoter. The predicted size of the fusion protein after signal peptide cleavage is 46 kDa. Live imaging of GFP demonstrates PV targeting of the fusion protein. Images are representative of two independent experiments. **b**, Western blot showing GFP^{DDD} is not degraded in the absence of TMP. Synchronized GFP^{DDD} parasites were grown 72 h with or without TMP before purification over Percoll and treatment with tetanolysin to release

the RBC cytosol but not the PV contents. The GFP^{DDD} fusion was detected with anti-HA antibodies. BiP serves as a loading control. Images represent one experiment. **c**, Growth analysis of asynchronous GFP^{DDD} parasites shows no growth defect -TMP. Error bars represent s.d. of three technical replicates. Data are representative of two independent experiments. **d**, IFA showing no defect in export in GFP^{DDD} parasites in the absence of TMP. No difference was observed in export of ring-specific (REX1) and the trophozoite-specific (MSRP6) exported proteins in the presence or absence of TMP. All scale bars, 5 μ m. Images represent one experiment.



Extended Data Figure 8 | Localization of PTEX components is unchanged under export blocking conditions. **a**, Immunoelectron microscopy showing localization of HSP101^{DDD} in ring-stage parasites with or without TMP. TMP was removed or not from synchronous late schizonts and parasites were allowed to re-invade and develop for 18 h before fixation. Scale bars, 500 nm. Images represent one experiment. **b**, **c**, IFA of 13F10 parasites showing co-localization between HSP101^{DDD} or EXP2 and the PV marker SERP with or without TMP, indicating that localization of these PTEX components is not

altered under export blocking conditions. TMP treatment was performed as in **a** before fixation with acetone and processing for IFA. **d**, **e**, IFA of ring-stage 13F10 parasites expressing a PTEX150–Flag fusion. The upper panel (**d**) shows co-localization with HSP101^{DDD}. The lower two panels (**e**) show that PTEX150 remains at the PVM during export block and partially colocalizes with blocked SBP1. TMP treatment and fixation as in **b**, **c**. All IFA scale bars, 5 μ m. All IFA images (**b**–**e**) are representative of two independent experiments.



Extended Data Figure 9 | Immunoprecipitation bead controls indicate target-specific interactions. Replicate IP experiments to those shown in Fig. 4c, e with bead controls included. **a, b**, Purification of RESA by HSP101^{DDD} or EXP2 (**a**) and purification of EXP2 by PTEX150^{Flag} (**b**) is specific to the target

antibodies. IP experiments were performed in parallel by incubating equivalent portions of lysate supernatant input with beads alone or beads and the indicated IP antibodies. S, supernatant input. E, elution. Data are representative of two independent experiments.

Equalizing excitation–inhibition ratios across visual cortical neurons

Mingshan Xue^{1,2,†}, Bassam V. Atallah³ & Massimo Scanziani^{1,2,4}

The relationship between synaptic excitation and inhibition (E/I ratio), two opposing forces in the mammalian cerebral cortex, affects many cortical functions such as feature selectivity and gain^{1,2}. Individual pyramidal cells show stable E/I ratios in time despite fluctuating cortical activity levels. This is because when excitation increases, inhibition increases proportionally through the increased recruitment of inhibitory neurons, a phenomenon referred to as excitation–inhibition balance^{3–9}. However, little is known about the distribution of E/I ratios across pyramidal cells. Through their highly divergent axons, inhibitory neurons indiscriminately contact most neighbouring pyramidal cells^{10,11}. Is inhibition homogeneously distributed¹² or is it individually matched to the different amounts of excitation received by distinct pyramidal cells? Here we discover that pyramidal cells in layer 2/3 of mouse primary visual cortex each receive inhibition in a similar proportion to their excitation. As a consequence, E/I ratios are equalized across pyramidal cells. This matched inhibition is mediated by parvalbumin-expressing but not somatostatin-expressing inhibitory cells and results from the independent adjustment of synapses originating from individual parvalbumin-expressing cells targeting different pyramidal cells. Furthermore, this match is activity-dependent as it is disrupted by perturbing pyramidal cell activity. Thus, the equalization of E/I ratios across pyramidal cells reveals an unexpected degree of order in the spatial distribution of synaptic strengths and indicates that the relationship between the cortex's two opposing forces is stabilized not only in time but also in space.

To determine the distribution of E/I ratios among layer 2/3 neighbouring pyramidal cells (Fig. 1a), we used adeno-associated virus (AAV) to conditionally express channelrhodopsin-2 (ChR2)^{13–15} in *Scnn1a-Cre-Tg3* mice and photoactivated layer 4 excitatory neurons, one of the main sources of synaptic excitation to layer 2/3, in acute visual cortical slices (Extended Data Fig. 1). We compared the E/I ratios between two to four simultaneously recorded layer 2/3 pyramidal cells (inter-soma distance $39.4 \pm 2.5 \mu\text{m}$, mean \pm s.e.m.; Extended Data Fig. 2) voltage clamped alternatively at the reversal potential for synaptic inhibition and excitation to isolate excitatory postsynaptic currents (EPSCs) and disynaptic inhibitory postsynaptic currents (IPSCs), respectively. EPSC amplitudes greatly varied between simultaneously recorded neurons and so did IPSC amplitudes (Fig. 1b). Despite the heterogeneous distributions of EPSC and IPSC amplitudes among pyramidal cells, however, we found a strong correlation between their amplitudes. That is, neurons with larger EPSCs also received larger IPSCs (Fig. 1c, e). As a consequence, the distribution of E/I ratios across pyramidal cells varied much less than the distributions of EPSC and IPSC amplitudes (Fig. 1d, f) and much less than if EPSCs and IPSCs were randomly paired between cells (Extended Data Fig. 2). These data indicate that E/I ratios are equalized across pyramidal cells.

This equalization could occur if distinct layer 2/3 pyramidal cells each receive inhibition from a ‘private’ set of inhibitory neurons such that the excitatory afferents that more strongly excite a pyramidal cell also more

strongly excite its private inhibitory neurons. However, the two classes of inhibitory neurons, parvalbumin-expressing (Pvalb) and somatostatin-expressing (Sst) cells, that provide most inhibition to layer 2/3 pyramidal

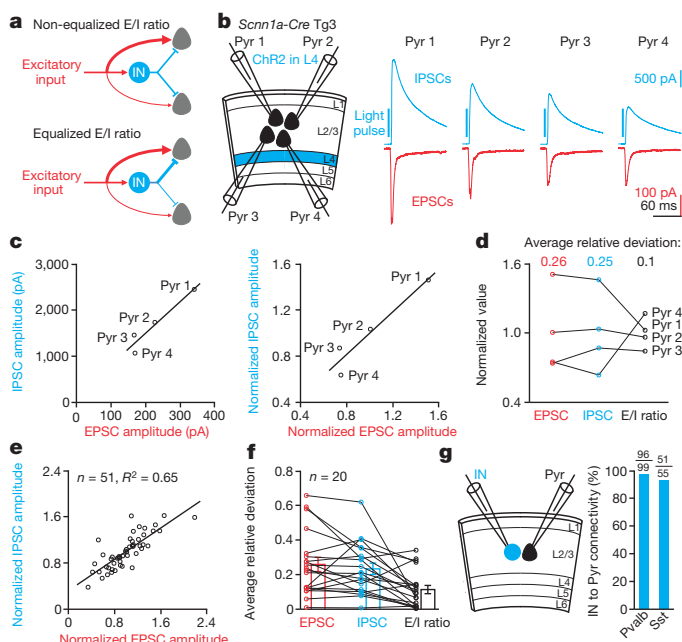


Figure 1 | Equalized E/I ratios across pyramidal cells. **a**, Two alternative models for the spatial distribution of E/I ratios across pyramidal cells. Top, the divergent axons of inhibitory neurons (IN) homogeneously inhibit neighbouring pyramidal cells. Pyramidal cells receiving more excitation have a larger E/I ratio. Bottom, despite divergent axons, inhibitory neurons generate larger inhibition in pyramidal cells receiving more excitation. Accordingly E/I ratios are equalized across pyramidal cells. **b**, Left, schematic of experiments. *Scnn1a-Cre-Tg3* mice with ChR2 in layer 4 excitatory neurons. Right, monosynaptic EPSCs and disynaptic IPSCs from four simultaneously recorded layer 2/3 pyramidal cells (Pyr) in response to layer 4 photoactivation. Note larger IPSCs in neurons receiving larger EPSCs. **c**, EPSC amplitudes of the four neurons in **b** plotted against their IPSC amplitudes. Left, absolute amplitudes. Right, normalized amplitudes. EPSC (or IPSC) amplitudes are normalized by the mean of the simultaneously recorded EPSC (or IPSC) amplitudes. Lines, linear regression fits. **d**, Distributions of normalized EPSC and IPSC amplitudes, and of normalized E/I ratios for the experiment in **b**. E/I ratios are normalized by the mean of the simultaneously recorded ratios. Note narrower distribution of E/I ratios compared with EPSCs or IPSCs. For computing average relative deviations, see Methods. **e**, Summary graphs of normalized EPSCs and IPSCs from 20 similar experiments ($n = 51$ cells). Line, linear regression fit ($R^2 = 0.65$, $P < 0.0001$). **f**, Summary graphs of average relative deviations from 20 similar experiments. Bars, mean \pm s.e.m. The average relative deviations of E/I ratios are 50% smaller than those of EPSCs ($P < 0.0001$) or IPSCs ($P < 0.0001$). **g**, Left, schematic of experiments. Right, connectivity rates from Pvalb and Sst cells to pyramidal cells.

¹Neurobiology Section, Division of Biological Sciences, Center for Neural Circuits and Behavior, University of California, San Diego, La Jolla, California 92093-0634, USA. ²Department of Neuroscience, University of California, San Diego, La Jolla, California 92093-0634, USA. ³Champalimaud Neuroscience Programme, Champalimaud Centre for the Unknown, Lisbon 1400-038, Portugal. ⁴Howard Hughes Medical Institute, University of California, San Diego, La Jolla, California 92093-0634, USA. [†]Present address: Department of Neuroscience, Baylor College of Medicine, Houston, Texas 77030, USA, and Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, Houston, Texas 77030, USA.

cells showed broad connectivity with pyramidal cells (97% and 93%, respectively, Fig. 1g), as previously shown^{10,11}, thus precluding the private connectivity.

Alternatively, the correlation between excitation and inhibition could be an artefact of the slicing procedure, whereby damaged neurons receive less excitation and less inhibition. To address this possibility we used an independent marker to identify neurons receiving more excitation. We used mice in which the promoter of the activity-dependent immediate early gene *Fos* drives the expression of *Fos* fused to the enhanced green fluorescent protein (*Fos-EGFP*), because in these mice $EGFP^+$ neurons receive more excitation than $EGFP^-$ neurons¹⁶. $EGFP^+$ neurons were predominantly pyramidal cells (Extended Data Fig. 3). We photo-stimulated layer 4 in acute slices from *Fos-EGFP*, *Scnn1a-Cre-Tg3* mice and simultaneously recorded pairs of $EGFP^+$ and nearby $EGFP^-$ layer 2/3 pyramidal cells. Layer 4 activation generated larger EPSCs in $EGFP^+$ neurons in 78% of all recorded pairs, and $EGFP^+$ neurons received, on average, 40% larger EPSCs (Fig. 2a, b) (the average logarithm of $EGFP^+/EGFP^-$ ratios was 0.15). Importantly, $EGFP^+$ neurons also received larger disynaptic IPSCs (Fig. 2a, c). Consequently, the E/I ratios of $EGFP^+$ and $EGFP^-$ neurons were similar (Fig. 2d).

Taken together, these results demonstrate that excitation and inhibition, despite varying in amplitudes between pyramidal cells, remain proportional to each other, thus equalizing E/I ratios.

Which type of interneuron provides the inhibition that matches layer-4-mediated excitation? We took advantage of the fact that $EGFP^+$ neurons in *Fos-EGFP* mice receive larger excitation from layer 4 and crossed them to *Pvalb-ires-Cre* or *Sst-ires-Cre* mice to express ChR2 conditionally. Photoactivation of Pvalb cells generated larger monosynaptic IPSCs in $EGFP^+$ than in $EGFP^-$ neurons (Fig. 2e, f). In contrast, Sst cells generated IPSCs whose amplitudes did not correlate with EGFP expression (Fig. 2g, h). These data indicate that Pvalb cells, but not Sst cells, provide stronger inhibition onto neurons that receive stronger layer-4-mediated excitation, thereby contributing to the equalization of E/I ratios.

What mechanism regulates the strengths of excitation and/or inhibition to achieve the observed proportionality? Excitation and inhibition may reach their specific ratio by using the pyramidal cell's activity as a measure of their relative strengths. For example, the low activity caused by a strong Pvalb-cell-mediated inhibition or by a weak layer-4-mediated excitation could be the signal to increase layer-4-mediated excitation or to decrease Pvalb-cell-mediated inhibition, respectively, until a neuron's specific higher set-point activity is reached. In both possibilities the initially small E/I ratio is increased by either increasing excitation to match the large inhibition or by decreasing inhibition to match the small excitation. Both possibilities are plausible since the activity of individual neurons can regulate the strengths of both excitatory and inhibitory synapses^{17–20}. If this hypothesis is correct, perturbing the activity of pyramidal cells should disrupt the proportionality between excitation and inhibition. For example, reducing the excitability of a pyramidal cell should increase its E/I ratio by either increasing excitation (the first possibility), or decreasing inhibition (the second possibility), or both.

We reduced the excitability of a small, random subset of layer 2/3 pyramidal cells in primary visual cortex (V1) by overexpressing a Kir2.1 channel via *in utero* electroporation (IUE)^{21–23} (Fig. 3a). Recordings in acute slices confirmed the reduced excitability in Kir2.1-overexpressing cells (Kir2.1 neurons) compared with untransfected control pyramidal cells (Extended Data Fig. 4). *In vivo* targeted recordings from Kir2.1 and nearby control neurons (Fig. 3b, c) demonstrated that Kir2.1 overexpression drastically suppressed visual-evoked and spontaneous activity (Fig. 3d–f). We then examined the impact of this perturbation on excitation and inhibition. We photostimulated layer 4 and simultaneously recorded Kir2.1 and neighbouring control neurons in the acute slices from *Scnn1a-Cre-Tg3* mice. Surprisingly, layer-4-mediated excitation was not significantly different between these two groups (Fig. 3g, h), invalidating the first aforementioned possibility. In contrast, disynaptic inhibition was significantly smaller in Kir2.1 neurons (Fig. 3g, i),

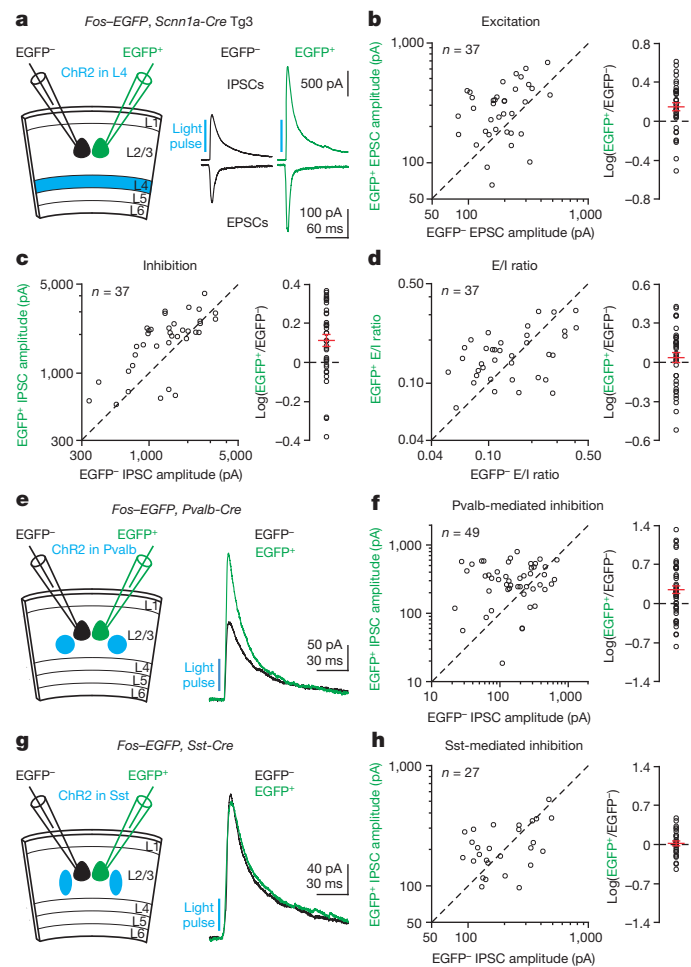


Figure 2 | Pvalb-cell-mediated inhibition matches layer-4-mediated excitation. **a**, Left, schematic of experiments. *Fos-EGFP*, *Scnn1a-Cre-Tg3* mice with ChR2 in layer 4 excitatory neurons. Right, monosynaptic EPSCs and disynaptic IPSCs from simultaneously recorded $EGFP^-$ and $EGFP^+$ neurons in response to layer 4 photoactivation. Note larger synaptic currents in $EGFP^+$ neuron. **b–d**, Summary graphs of 37 similar experiments. **b**, Left, EPSC amplitudes in $EGFP^+$ neurons plotted against those in $EGFP^-$ neurons. Right, logarithm of the ratio between EPSC amplitudes in $EGFP^+$ and $EGFP^-$ neurons. Red, mean ± s.e.m. EPSC amplitudes are 40% larger in $EGFP^+$ neurons ($P = 0.0004$). **c**, As in **b**, but for disynaptic IPSCs. Disynaptic IPSC amplitudes are 30% larger in $EGFP^+$ neurons ($P = 0.001$). **d**, As in **b**, but for E/I ratios. E/I ratios are similar between $EGFP^+$ and $EGFP^-$ neurons ($P = 0.7$). **e**, Left, schematic of experiments. *Fos-EGFP*, *Pvalb-ires-Cre* mice with ChR2 in Pvalb cells. Right, IPSCs from simultaneously recorded $EGFP^-$ and $EGFP^+$ neurons in response to Pvalb cell photoactivation. Note larger IPSC in $EGFP^+$ neuron. **f**, Summary graph. Left, IPSC amplitudes in $EGFP^+$ neurons plotted against those in $EGFP^-$ neurons. Right, logarithm of the ratio between IPSC amplitudes in $EGFP^+$ and $EGFP^-$ neurons. Red, mean ± s.e.m. IPSC amplitudes are 77% larger in $EGFP^+$ neurons ($n = 49$, $P = 0.001$). **g**, **h**, As in **e**, **f**, but for *Fos-EGFP*, *Sst-ires-Cre* mice with ChR2 in Sst cells. IPSC average amplitudes are similar between $EGFP^+$ and $EGFP^-$ neurons ($n = 27$, $P = 0.7$).

consistent with the second possibility. The effect on inhibition was due to the channel function of Kir2.1 because a non-conducting Kir2.1 mutant (Extended Data Fig. 4) had no effect (Extended Data Fig. 5). Thus, perturbing layer 2/3 pyramidal cell excitability disrupts the proportionality between excitation and inhibition (Fig. 3j). These data indicate that pyramidal cell activity contributes to the equalization of E/I ratios across pyramidal cells.

If pyramidal cell activity contributes to establishing the proportionality between layer-4-mediated excitation and Pvalb-cell-mediated inhibition, then the decrease in excitability should selectively decrease

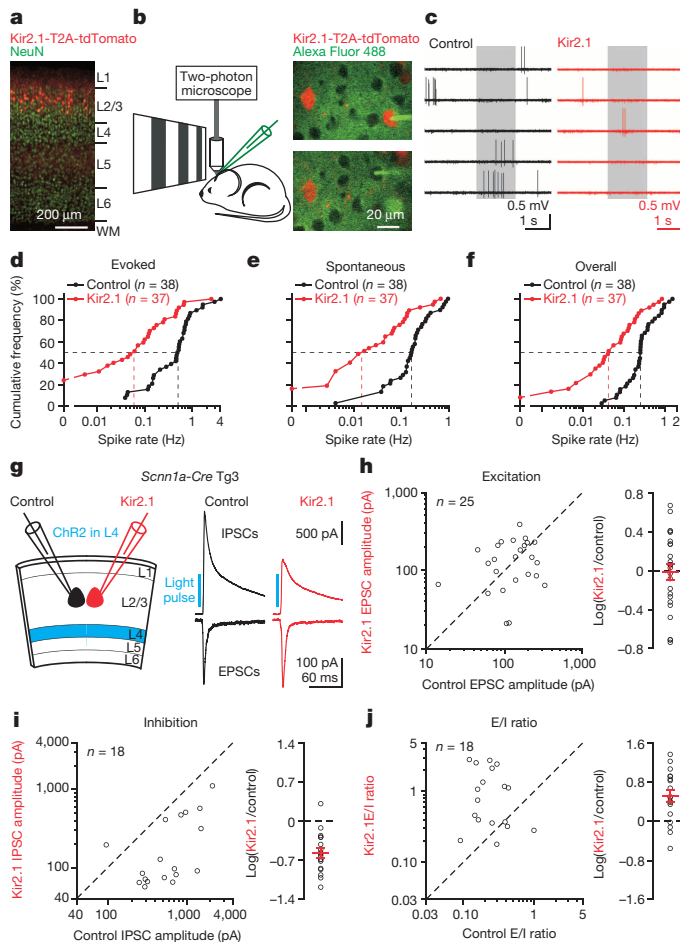


Figure 3 | Suppressing pyramidal cell activity reduces inhibition but not excitation. **a**, Fluorescent image of a V1 coronal section showing Kir2.1-T2A-TdTomato overexpression in a small subset of layer 2/3 pyramidal cells ($9 \pm 1\%$, mean \pm s.e.m., $n = 12$ sections from six mice). Cortical layers are identified by NeuN staining. L, layer, WM, white matter. **b**, Left, schematic of *in vivo* experiments. Right, a Kir2.1 neuron (upper panel) and a control neuron (lower panel) were sequentially recorded with Alexa Fluor 488-filled pipettes. **c**, Recordings from a control and a Kir2.1 neuron show spontaneous and visual-evoked spikes. Grey box, visual stimulation period. Note reduced spiking in Kir2.1 neuron. **d–f**, Cumulative frequencies of evoked spike rate (**d**, median: control, 0.50 Hz; Kir2.1, 0.061 Hz; $P < 0.0001$), spontaneous spike rate (**e**, median: control, 0.16 Hz; Kir2.1, 0.017 Hz; $P < 0.0001$) and overall spike rate (**f**, median: control, 0.25 Hz; Kir2.1, 0.043 Hz; $P < 0.0001$) from 38 control neurons and 37 Kir2.1 neurons. **g**, Left, schematic of slice experiments. *Scnn1a-Cre-Tg3* mice with ChR2 in layer 4 excitatory neurons and Kir2.1 in a subset of layer 2/3 pyramidal cells. Right, monosynaptic EPSCs and disynaptic IPSCs from simultaneously recorded control and Kir2.1 neurons in response to layer 4 photoactivation. Note similar EPSC but smaller disynaptic IPSC in Kir2.1 neuron compared with control neuron. **h–j**, Summary graphs. **h**, Left, EPSC amplitudes in Kir2.1 neurons plotted against those in control neurons. Right, logarithm of the ratio between EPSC amplitudes in Kir2.1 and control neurons. Red, mean \pm s.e.m. EPSC average amplitudes are similar between Kir2.1 and control neurons ($n = 25$, $P = 0.8$). **i**, As in **h**, but for disynaptic IPSCs. Disynaptic IPSC amplitudes in Kir2.1 neurons are 27% of those in control neurons ($n = 18$, $P = 0.0003$). **j**, As in **h**, but for E/I ratios. E/I ratios in Kir2.1 neurons are threefold those in control neurons ($n = 18$, $P = 0.004$).

Pvalb- but not Sst-cell-mediated inhibition. Conversely, an increase in excitability should selectively increase Pvalb-cell-mediated inhibition. Indeed, Pvalb-cell-mediated inhibition was significantly smaller in Kir2.1 than in control neurons, whereas Sst-cell-mediated inhibition was similar (Fig. 4a–e). Overexpression of the non-conducting Kir2.1 mutant did not affect Pvalb-cell-mediated inhibition (Extended Data Fig. 5). We used a bacterial voltage-gated Na^+ channel (mNaChBac) to enhance neuronal excitability. Neurons expressing mNaChBac generate

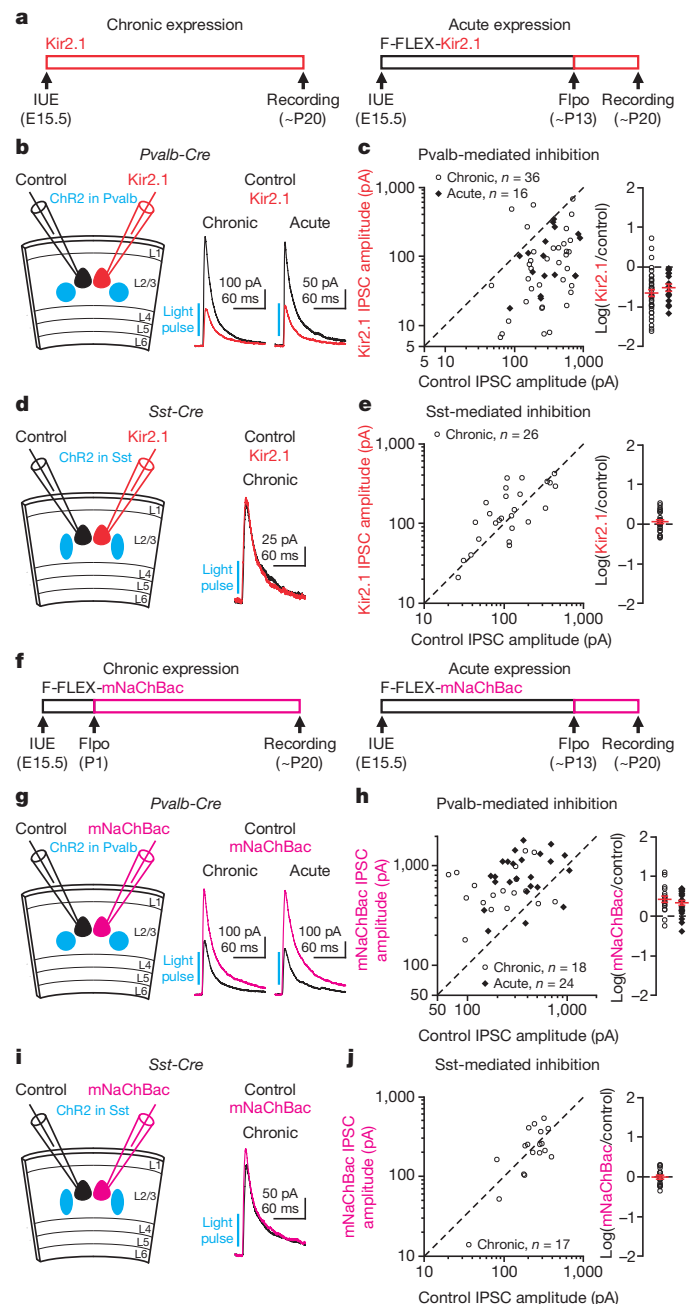


Figure 4 | Bidirectional regulation of Pvalb- but not Sst-cell-mediated inhibition. **a**, Schematic of chronic and acute expression of Kir2.1. Red bars, approximate Kir2.1 expression time course. **b**, Left, schematic of experiments. *Pvalb-ires-Cre* mice with ChR2 in Pvalb cells and Kir2.1 in a subset of layer 2/3 pyramidal cells. Right, IPSCs from simultaneously recorded control neuron and neuron chronically or acutely expressing Kir2.1 in response to Pvalb cell photoactivation. Note smaller IPSCs in Kir2.1 neuron. **c**, Summary graphs. Left, IPSC amplitudes in Kir2.1 neurons plotted against those in control neurons. Right, logarithm of the ratio between IPSC amplitudes in Kir2.1 and control neurons. Red, mean \pm s.e.m. IPSC amplitudes in Kir2.1 neurons are 23% ($n = 36$, $P < 0.0001$) and 31% ($n = 16$, $P = 0.0005$) of those in control neurons for chronic and acute conditions, respectively. **d**, **e**, As in **b**, **c**, but for *Sst-ires-Cre* mice with ChR2 in *Sst* cells and Kir2.1 chronically in a subset of layer 2/3 pyramidal cells. On average IPSC amplitudes are similar between Kir2.1 and control neurons ($n = 26$, $P = 0.3$). **f**, Schematic of chronic and acute expression of mNaChBac. Magenta bars, approximate mNaChBac expression time course. **g**, **h**, As in **b**, **c**, but for mNaChBac. IPSC amplitudes in mNaChBac neurons are 2.7-fold ($n = 18$, $P = 0.001$) and 2.2-fold ($n = 24$, $P = 0.0003$) those in control neurons for chronic and acute conditions, respectively. **i**, **j**, As in **d**, **e**, but for mNaChBac. On average, IPSC amplitudes are similar between mNaChBac and control neurons ($n = 17$, $P = 0.7$).

long-lasting action potentials and depolarization of the order of hundreds of milliseconds (Extended Data Fig. 6). Because constitutive expression of mNaChBac in cortical neurons from embryonic day 15.5 (E15.5) caused a neuronal migration defect (Extended Data Fig. 7), we devised a Flpo recombinase-mediated flip-excision strategy, F-FLEX switch (Extended Data Fig. 8), to conditionally express mNaChBac postnatally. We combined *in utero* electroporation of a Flpo-dependent mNaChBac-expressing plasmid, to randomly transfect a small subset of layer 2/3 pyramidal cells, with injection of an AAV expressing Flpo at postnatal day 1 (P1), to turn on mNaChBac expression. This allowed us to concurrently express ChR2 in Pvalb or Sst cells, and mNaChBac in layer 2/3 pyramidal cells without affecting their migration (Extended Data Fig. 7). Pvalb-cell-mediated inhibition was significantly larger in mNaChBac neurons than in control neurons (Fig. 4f–h), and a non-conducting mNaChBac mutant (Extended Data Fig. 6) had no effect (Extended Data Fig. 5). mNaChBac expression did not alter Sst-cell-mediated inhibition (Fig. 4i, j). To determine whether also more acute perturbations of layer 2/3 pyramidal cell excitability alter Pvalb-cell-mediated inhibition, we used Flpo and F-FLEX switch to express Kir2.1 or mNaChBac for only approximately 1 week starting around postnatal days 12–14. This acute decrease (Kir2.1) or increase (mNaChBac) in excitability caused a decrease or an increase in Pvalb-cell-mediated inhibition, respectively, similar to the changes caused by the chronic expression of Kir2.1 or mNaChBac (Fig. 4a–c, f–h). These data indicate that the proportionality between layer-4-mediated excitation and Pvalb-cell-mediated inhibition is equalized across pyramidal cells through the bidirectional modulation of the strength of Pvalb cell synapses.

The above results show that the spatial heterogeneity of Pvalb-cell-mediated inhibition ensures the equalization of E/I ratios across pyramidal cells. Is the inhibition mediated by a single Pvalb cell also heterogeneous across its targeted pyramidal cells? We first determined whether the relative amplitudes of unitary IPSCs (uIPSCs) mediated by a Pvalb cell onto its targets are predicted by the relative activity of these targets. We suppressed the activity of a small subset of layer 2/3 pyramidal cells by overexpressing Kir2.1 and simultaneously recorded from a layer 2/3 Pvalb cell, a control and a Kir2.1 neuron (Fig. 5a). Although the Pvalb-to-pyramidal cell connectivity was similarly high, regardless of whether pyramidal cells overexpressed Kir2.1 (Fig. 5f), uIPSC amplitude was significantly smaller in Kir2.1 neurons than in control neurons (Fig. 5b, f). All properties of the unitary connections between Pvalb cells and control neurons were similar to those recorded in mice that were not transfected with Kir2.1 (Extended Data Fig. 9), indicating a cell-autonomous effect of Kir2.1 overexpression. We assessed the variability of uIPSC amplitudes originating from a single Pvalb cell and determined its dependency on the activity of the targeted pyramidal cells. We simultaneously recorded from a layer 2/3 Pvalb cell and two or three nearby pyramidal cells that were either all control or all Kir2.1 neurons (Fig. 5c, d). uIPSC amplitudes varied greatly from one control neuron to another, but less among Kir2.1 neurons (Fig. 5c–e, g), possibly because suppressing pyramidal cell activity cannot reduce uIPSC amplitudes below a certain level (flooring effect) (Extended Data Fig. 10). Thus, the inhibition generated by even an individual Pvalb cell onto its targets is remarkably heterogeneous, and this heterogeneity reflects in part the activity profile of the targeted pyramidal cell population. Hence, despite the indiscriminate connectivity of Pvalb cells, the amount of inhibition that they provide onto each of their targets is adjusted to equalize the E/I ratios (Fig. 5h).

Both theoretical and experimental evidence indicates that the relationship between synaptic excitation and inhibition in the cerebral cortex is fundamental for sensory processing^{1,2,24–27}. Failure to establish or maintain this relationship may be the neural basis of neurological disorders such as schizophrenia and autism^{28,29}. We discover that E/I ratios are remarkably similar across different pyramidal cells despite large variations in the amplitudes of synaptic excitation and inhibition. The activity-dependent adjustment of inhibition to match excitation may result from activity-dependent gene expression^{19,20,30}. Our study provides insight into how two opposing synaptic inputs, layer-4-mediated

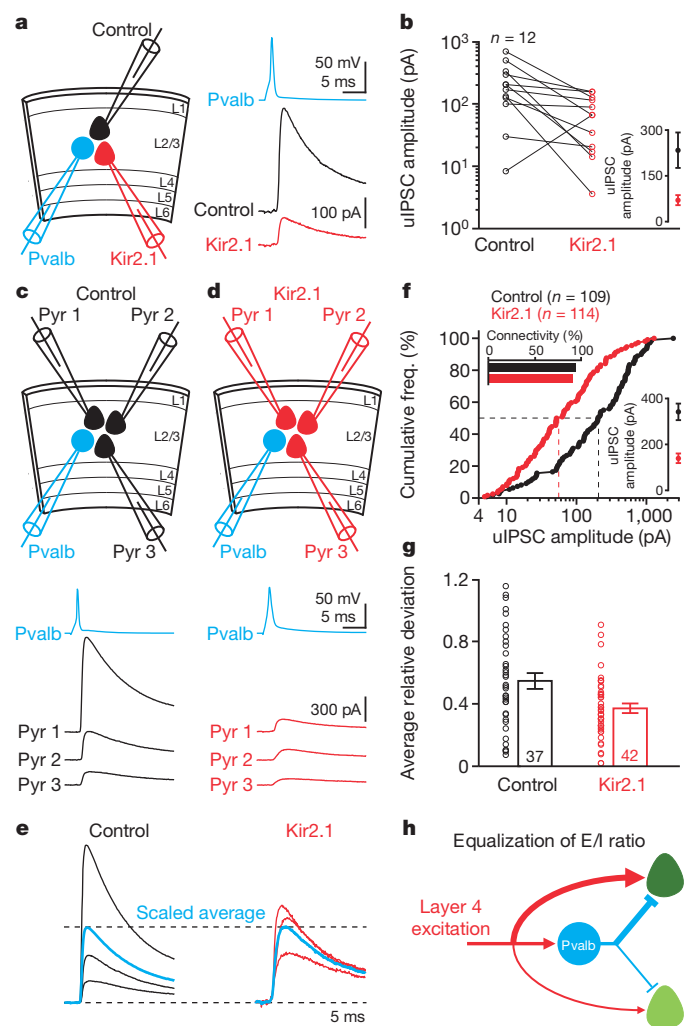


Figure 5 | Inhibition mediated by individual Pvalb cells varies depending on targets' activity. **a**, Left, schematic of experiments. Right, uIPSCs from simultaneously recorded control and Kir2.1 neuron in response to an action potential in a Pvalb cell. Note smaller uIPSC in Kir2.1 neuron. **b**, Summary graphs of uIPSC amplitudes from 12 similar experiments ($P = 0.005$). Inset, mean \pm s.e.m. **c**, Top, schematic of experiments. Bottom, uIPSCs simultaneously recorded from three control neurons in response to an action potential in a Pvalb cell. **d**, As in **c**, but for three Kir2.1 neurons. **e**, The uIPSCs from **c** and **d** were normalized by their respective mean amplitudes (scaled average). Note larger inter-cell variability of uIPSCs among control neurons. **f**, Cumulative frequencies of uIPSC amplitudes (control: $n = 109$, median, 205.4 pA; Kir2.1: $n = 114$, 57.4 pA; $P < 0.0001$). Lower inset, mean \pm s.e.m. of uIPSC amplitudes. Upper inset, unitary connectivity rates from Pvalb cells to control (109 out of 116) and Kir2.1 (114 out of 126) neurons are similar ($P = 0.3$). **g**, Summary graph for the average relative deviations of uIPSCs from 37 and 42 experiments as in **c** and **d**. Bars, mean \pm s.e.m. The average relative deviations for Kir2.1 neurons are 33% smaller than those for control neurons ($P = 0.009$). **h**, Schematic of the equalized E/I ratios across cortical neurons.

excitation and Pvalb-cell-mediated inhibition, remain proportional across a population of pyramidal cells. Thus it reveals an unanticipated degree of order in the distribution of synaptic strengths in cortical space.

METHODS SUMMARY

Cre-dependent recombinant AAV vectors were injected postnatally into Cre-expressing mice to conditionally express ChR2. Plasmids were electroporated *in utero* at embryonic day 14.5 or 15.5 to transfect a small, random subset of layer 2/3 pyramidal cells. For *in vitro* physiology, mice at postnatal days 14–23 were anaesthetized and transcardially perfused. Coronal slices were perfused with artificial cerebrospinal fluid for whole-cell recordings at 31–32 °C. A light-emitting diode (470 nm) was used to deliver blue light to stimulate neurons via the activation of

ChR2. For *in vivo* physiology, mice at postnatal days 17–23 were anaesthetized by intraperitoneal injection of chlorprothixene (5 mg kg^{-1}) followed by urethane (1.2 g kg^{-1}) and the body temperature was maintained at 37°C . A craniotomy at V1 was performed and targeted loose-patch recordings were performed under the guidance of two-photon laser scanning microscopy.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 2 January; accepted 8 April 2014.

Published online 22 June 2014.

- Haider, B. & McCormick, D. A. Rapid neocortical dynamics: cellular and network mechanisms. *Neuron* **62**, 171–189 (2009).
- Isaacson, J. S. & Scanziani, M. How inhibition shapes cortical activity. *Neuron* **72**, 231–243 (2011).
- Anderson, J. S., Carandini, M. & Ferster, D. Orientation tuning of input conductance, excitation, and inhibition in cat primary visual cortex. *J. Neurophysiol.* **84**, 909–926 (2000).
- Shu, Y., Hasenstaub, A. & McCormick, D. A. Turning on and off recurrent balanced cortical activity. *Nature* **423**, 288–293 (2003).
- Wehr, M. & Zador, A. M. Balanced inhibition underlies tuning and sharpens spike timing in auditory cortex. *Nature* **426**, 442–446 (2003).
- Wilent, W. B. & Contreras, D. Synaptic responses to whisker deflections in rat barrel cortex as a function of cortical layer and stimulus intensity. *J. Neurosci.* **24**, 3985–3998 (2004).
- Haider, B., Duque, A., Hasenstaub, A. R. & McCormick, D. A. Neocortical network activity *in vivo* is generated through a dynamic balance of excitation and inhibition. *J. Neurosci.* **26**, 4535–4545 (2006).
- Okun, M. & Lampl, I. Instantaneous correlation of excitation and inhibition during ongoing and sensory-evoked activities. *Nature Neurosci.* **11**, 535–537 (2008).
- Atallah, B. V. & Scanziani, M. Instantaneous modulation of gamma oscillation frequency by balancing excitation with inhibition. *Neuron* **62**, 566–577 (2009).
- Fino, E. & Yuste, R. Dense inhibitory connectivity in neocortex. *Neuron* **69**, 1188–1203 (2011).
- Packer, A. M. & Yuste, R. Dense, unspecific connectivity of neocortical parvalbumin-positive interneurons: a canonical microcircuit for inhibition? *J. Neurosci.* **31**, 13260–13271 (2011).
- Pouille, F., Marin-Burgin, A., Adesnik, H., Atallah, B. V. & Scanziani, M. Input normalization by global feedforward inhibition expands cortical dynamic range. *Nature Neurosci.* **12**, 1577–1585 (2009).
- Nagel, G. *et al.* Channelrhodopsin-2, a directly light-gated cation-selective membrane channel. *Proc. Natl Acad. Sci. USA* **100**, 13940–13945 (2003).
- Boyden, E. S., Zhang, F., Bamberg, E., Nagel, G. & Deisseroth, K. Millisecond-timescale, genetically targeted optical control of neural activity. *Nature Neurosci.* **8**, 1263–1268 (2005).
- Li, X. *et al.* Fast noninvasive activation and inhibition of neural and network activity by vertebrate rhodopsin and green algae channelrhodopsin. *Proc. Natl Acad. Sci. USA* **102**, 17816–17821 (2005).
- Benedetti, B. L., Takashima, Y., Wen, J. A., Urban-Ciecko, J. & Barth, A. L. Differential wiring of layer 2/3 neurons drives sparse and reliable firing during neocortical development. *Cereb. Cortex* **23**, 2690–2699 (2013).
- Burrone, J., O'Byrne, M. & Murthy, V. N. Multiple forms of synaptic plasticity triggered by selective suppression of activity in individual neurons. *Nature* **420**, 414–418 (2002).
- Ibata, K., Sun, Q. & Turrigiano, G. G. Rapid synaptic scaling induced by changes in postsynaptic firing. *Neuron* **57**, 819–826 (2008).
- Peng, Y. R. *et al.* Postsynaptic spiking homeostatically induces cell-autonomous regulation of inhibitory inputs via retrograde signaling. *J. Neurosci.* **30**, 16220–16231 (2010).
- Sim, S., Antolin, S., Lin, C. W., Lin, Y. X. & Lois, C. Increased cell-intrinsic excitability induces synaptic changes in new neurons in the adult dentate gyrus that require npas4. *J. Neurosci.* **33**, 7928–7940 (2013).
- Saito, T. & Nakatsuji, N. Efficient gene transfer into the embryonic mouse brain using *in vivo* electroporation. *Dev. Biol.* **240**, 237–246 (2001).
- Tabata, H. & Nakajima, K. Efficient *in utero* gene transfer system to the developing mouse brain using electroporation: visualization of neuronal migration in the developing cortex. *Neuroscience* **103**, 865–872 (2001).
- Hatanaka, Y., Hisanaga, S., Heizmann, C. W. & Murakami, F. Distinct migratory behavior of early- and late-born neurons derived from the cortical ventricular zone. *J. Comp. Neurol.* **479**, 1–14 (2004).
- van Vreeswijk, C. & Sompolinsky, H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* **274**, 1724–1726 (1996).
- Vogels, T. P. & Abbott, L. F. Gating multiple signals through detailed balance of excitation and inhibition in spiking networks. *Nature Neurosci.* **12**, 483–491 (2009).
- Renart, A. *et al.* The asynchronous state in cortical circuits. *Science* **327**, 587–590 (2010).
- Vogels, T. P., Sprekeler, H., Zenke, F., Clopath, C. & Gerstner, W. Inhibitory plasticity balances excitation and inhibition in sensory pathways and memory networks. *Science* **334**, 1569–1573 (2011).
- Rubenstein, J. L. & Merzenich, M. M. Model of autism: increased ratio of excitation/inhibition in key neural systems. *Genes Brain Behav.* **2**, 255–267 (2003).
- Lewis, D. A., Curley, A. A., Glausier, J. R. & Volk, D. W. Cortical parvalbumin interneurons and cognitive dysfunction in schizophrenia. *Trends Neurosci.* **35**, 57–67 (2012).
- Bloodgood, B. L., Sharma, N., Browne, H. A., Trepman, A. Z. & Greenberg, M. E. The activity-dependent transcription factor NPAS4 regulates domain-specific inhibition. *Nature* **503**, 121–125 (2013).

Acknowledgements We thank M. Chan, J. Evora, A. Linder and P. Abelkop for technical assistance; M. S. Caudill and S. R. Olsen for help with the *in vivo* physiology recording programme; E. Kim and A. Ghosh for pCAG-Kir2.1-T2A-tdTomato plasmid; J. Isaacson and H. Y. Zoghbi for comments on earlier versions of the manuscript; D. N. Hill, G. I. Allen, E. Arias-Castro and M. Wang for advice on statistical analysis; the members of the Scanziani and Isaacson laboratories for suggestions; and the University of California, San Diego Neuroscience Microscopy Facility (P30 NS047101) for imaging equipment. M.X. was supported by a fellowship from Jane Coffin Childs Memorial Fund for Medical Research. M.S. is an investigator of the Howard Hughes Medical Institute. This work was also supported by the Gatsby Charitable Foundation.

Author Contributions M.X. and M.S. designed the study. M.X. performed all experiments and data analysis. B.V.A. contributed to data analysis. M.X. and M.S. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.X. (mingshanxue@gmail.com) or M.S. (massimo@ucsd.edu).

METHODS

Mice. All procedures to maintain and use mice were approved by the Institutional Animal Care and Use Committee at the University of California, San Diego. Mice were maintained on a reverse 12-h:12-h light:dark cycle with regular mouse chow and water *ad libitum*. CD-1 mice were purchased from Charles River Laboratories or Harlan Laboratories. *Scnn1a-Cre-Tg3* (ref. 31), *Fos-EGFP*³², *Gad2-ires-Cre*³³, *Pvalb-ires-Cre*³⁴, *Sst-ires-Cre*³⁵ and *Rosa-CAG-LSL-tdTomato-WPRE*³¹ mice were obtained from the Jackson Laboratory (stock numbers 009613, 014135, 010802, 008069, 013044 and 007909, respectively). Hemizygous transgenic mice and heterozygous knock-in mice of both sexes were used in the experiments.

DNA constructs and transfection of HEK cells. Two point mutations E224G and Y242F were introduced into mouse wild-type Kir2.1 (Kcnj2) to enhance its ability to suppress neuronal activity. Mutation E224G attenuates the Mg²⁺ and polyamine block of Kir2.1 to reduce its inward rectification³⁵. Mutation Y242F blocks tyrosine kinase phosphorylation of Kir2.1 at residue Y242 to enhance its plasma membrane surface expression³⁶. Three point mutations, G144A, Y145A and G146A, were introduced to generate a non-conducting channel³⁷. A Myc tag (EQKLISEEDL) was fused to the amino (N) termini of Kir2.1 E224G Y242F and Kir2.1 E224G Y242F G144A Y145A G146A, referred to as Kir2.1 and Kir2.1Mut, respectively. Both Kir2.1 and Kir2.1Mut were carboxy (C)-terminally fused with a T2A sequence (GSGEGRSLTTCGDVEENPGP) followed by a tdTomato. The resulting constructs were then cloned into a plasmid containing a CAG promoter (pCAG) to generate pCAG-Kir2.1-T2A-tdTomato and pCAG-Kir2.1Mut-T2A-tdTomato.

The complementary DNA (cDNA) encoding a wild-type bacterial Na⁺ channel NaChBac³⁸ was synthesized *de novo* and codon-optimized for mammalian expression (referred as mNaChBac) by DNA2.0. A point mutation E191K was introduced to generate a non-conducting channel³⁹, referred as mNaChBacMut. Both mNaChBac and mNaChBacMut were C-terminally fused with T2A-tdTomato and cloned into the pCAG plasmid to create pCAG-mNaChBac-T2A-tdTomato and pCAG-mNaChBacMut-T2A-tdTomato, respectively.

F-FLEX cassette using two wild-type Frt sites and two F14 sites⁴⁰ (Extended Data Fig. 8) was synthesized *de novo* and cloned into the plasmid pJ244 by DNA2.0 to generate pJ244-F-FLEX. mNaChBac-T2A-tdTomato and mNaChBacMut-T2A-tdTomato were subcloned into pJ244-F-FLEX in the inverted orientation. F-FLEX-mNaChBac-T2A-tdTomato and F-FLEX-mNaChBacMut-T2A-tdTomato cassettes were then subcloned into an AAV *cis*-plasmid containing an EF1 α promoter to generate pAAV-EF1 α -F-FLEX-mNaChBac-T2A-tdTomato and pAAV-EF1 α -F-FLEX-mNaChBacMut-T2A-tdTomato, respectively. mNaChBac-T2A-tdTomato in pAAV-EF1 α -F-FLEX-mNaChBac-T2A-tdTomato was replaced with inverted Kir2.1-T2A-tdTomato to generate pAAV-EF1 α -F-FLEX-Kir2.1-T2A-tdTomato.

An improved version of Flp recombinase, Flpo⁴¹ was cloned into a pCAG plasmid and an AAV *cis*-plasmid containing a human synapsin promoter to generate pCAG-Flpo and pAAV-hSynapsin-Flpo, respectively. pCAG-EGFP⁴², pCAG-mRFP⁴³ and pCAG-Cre⁴⁴ were obtained from Addgene (11150, 28311 and 13775, respectively).

HEK-293FT cells (Life Technologies) were transfected with DNA constructs (0.1–0.5 μ g) in 12-well plates using Lipofectamine 2000 (Life Technologies) to test their functionality. The following constructs were used in Extended Data Fig. 8c, d: pCAG-EGFP, pCAG-mRFP, pCAG-Flpo, pCAG-Cre, pAAV-EF1 α -F-FLEX-mNaChBac-T2A-tdTomato and pAAV-EF1 α -DIO-hChR2(H134R)-EYFP.

In utero electroporation. Female CD-1 mice were crossed with male *Scnn1a-Cre-Tg3*, *Pvalb-ires-Cre*, or *Sst-ires-Cre* mice to obtain timed pregnancies. pCAG-Kir2.1-T2A-tdTomato, pCAG-Kir2.1Mut-T2A-tdTomato, pCAG-mNaChBac-T2A-tdTomato, and pCAG-mNaChBacMut-T2A-tdTomato were used at the final concentrations of 2–3 μ g μ l⁻¹. pAAV-EF1 α -F-FLEX-mNaChBac-T2A-tdTomato, pAAV-EF1 α -F-FLEX-mNaChBacMut-T2A-tdTomato or pAAV-EF1 α -F-FLEX-Kir2.1-T2A-tdTomato (2–3 μ g μ l⁻¹ final concentration) was mixed with pCAG-EGFP (0.2 μ g μ l⁻¹ final concentration). Fast Green (Sigma-Aldrich, 0.01% final concentration) was added to the DNA solution. On embryonic day 14.5 or 15.5, female mice were anaesthetized with 2.5% isoflurane in oxygen at a flow rate of 1 l min⁻¹ and the body temperature was maintained by a feedback-based DC temperature control system (FHC) at 37 °C. Buprenorphine (3 μ g, Reckitt Benckiser Healthcare) was administered subcutaneously along with 1 ml of Lactated Ringer's Injection (Baxter Healthcare). The abdominal fur was shaved and the skin was cleaned with 70% alcohol and iodine. A sterile towel drape was laid on the mouse with only the abdomen exposed. Midline incisions (2 cm) were made on the abdominal skin and wall, and the uteri were taken out of the abdominal cavity. A bevelled glass micropipette (tip size 100- μ m outer diameter, 50- μ m inner diameter) was used to penetrate the uterus and the embryo skull to inject about 1.5 μ l of DNA solution into one lateral ventricle. Five pulses of current (voltage 39 V, duration 50 ms) were delivered at 1 Hz with a Tweezertrode (5-mm diameter) and a square-wave pulse generator (ECM 830, BTX Harvard Apparatus). The electrode paddles were positioned along the 70° angle to the brain's sagittal plane. The cathode faced the occipital side of the injected ventricle to target the visual cortex. After electroporation, uteri were put back into the

abdominal cavity, and the abdominal wall and skin were sutured. Mice were returned to the home cage and recovered from anaesthesia on a 37 °C Deltaphase Isothermal Pad (Braintree Scientific). Additional buprenorphine (3 μ g) was administered subcutaneously on the next day. After birth, transfected pups were identified by the transcranial fluorescence of tdTomato or EGFP with a stereomicroscope (MVX10 Macroview, Olympus). Only the pups in which the majority of the transfection occurred in the primary visual cortex were used for experiments.

AAV production and injection. All recombinant AAV vectors were produced by the Penn Vector Core with the following titres: AAV2/9-CAGGS-Flex-ChR2-tdTomato⁴⁵ (Addgene 18917, titre 1.15 \times 10¹³ genome copies per millilitre), AAV2/1-CAGGS-Flex-ChR2-tdTomato (titre 6.86 \times 10¹² or 1.22 \times 10¹³ genome copies per millilitre), AAV2/9-EF1 α -DIO-hChR2(H134R)-EYFP⁴⁶ (Addgene 20298, titre 6.24 \times 10¹² or 1.18 \times 10¹³ genome copies per millilitre), AAV2/1-EF1 α -DIO-hChR2(H134R)-EYFP (titre 3.41 \times 10¹³ genome copies per millilitre) and AAV2/9-hSynapsin-Flpo (titre 1.57 \times 10¹³ genome copies per millilitre).

Injection at postnatal days 0–2. AAV was injected into the V1 of pups using a Nanoject II nanolitre injector (Drummond Scientific Company). Pups were anaesthetized by hypothermia and secured on a custom-made plate. Fast Green (0.01% final concentration) was added to the virus solution for visualization. A bevelled glass micropipette (tip size 60- μ m outer diameter, 30- μ m inner diameter) was used to penetrate the scalp and skull, and to inject AAV at different depths (600, 500, 400 and 300 μ m below the scalp) of one location (1.6 mm lateral and 0.3 mm caudal from the lambda). A total of about 80–180 nl (adjusted based on the virus titres) of virus solution was injected over 60 s. After injection, the micropipette was kept in the parenchyma at 300- μ m depth for 30 s before being slowly withdrawn. Pups were placed on a 37 °C Deltaphase Isothermal Pad to recover from anaesthesia and then were returned to the dam. For chronic expression of mNaChBac or mNaChBacMut, a mix of AAV2/9-hSynapsin-Flpo (titre 1.57 \times 10¹³ genome copies per millilitre) and AAV2/9-EF1 α -DIO-hChR2(H134R)-EYFP (titre 1.18 \times 10¹³ genome copies per millilitre) at a ratio of 1:3 was injected on postnatal day 1.

Injection at postnatal days 12–14. For acute expression of Kir2.1 or mNaChBac, mice previously injected with AAV2/1-EF1 α -DIO-hChR2(H134R)-EYFP between postnatal days 0 and 2 were injected with AAV2/9-hSynapsin-Flpo between postnatal days 12 and 14. Mice were anaesthetized with 2.5% isoflurane in oxygen at a flow rate of 1 l min⁻¹ and the body temperature was maintained by a feedback-based DC temperature control system at 37 °C. Buprenorphine (1 μ g) was administered subcutaneously along with 0.1 ml of Lactated Ringer's Injection. Lubricant ophthalmic ointment (Artificial Tears Ointment, Rugby Laboratories) was applied to the corneas to prevent drying. The scalp fur was shaved and the skin was cleaned with 70% alcohol and iodine. A small incision (0.5 cm) was made on the skin medial to the visual cortex. The skull at the injection site (2.5 mm lateral to the midline and 1 mm rostral to the lambda suture; the same site that was previously electroporated *in utero* and virally injected between postnatal days 0 and 2) was thinned with a 0.3-mm diameter round bur (Busch Bur, Gesswein) on a high-speed rotary micro-motor (Foredom) such that the injection glass micropipette (tip size 50- μ m outer diameter, 25- μ m inner diameter) could penetrate the skull. A total of 150 nl of virus solution was injected 450 μ m below the skull at a rate of 20 nl min⁻¹ using an UltraMicroPump III and a Micro4 controller (World Precision Instruments). After the injection, the micropipette was kept in the parenchyma for 5–10 min before being slowly withdrawn. The skin was sutured. Mice were returned to their home cage to recover from anaesthesia on a 37 °C Deltaphase Isothermal Pad.

Immunocytochemistry. Mice were anaesthetized by an intraperitoneal injection of a ketamine and xylazine mix (100 mg kg⁻¹ and 10 mg kg⁻¹, respectively), and were transcardially perfused with phosphate buffered saline (PBS, pH 7.4) followed by 4% paraformaldehyde in PBS (pH 7.4). Brains were removed, further fixed overnight in 4% paraformaldehyde, cryoprotected with 30% sucrose in PBS and frozen in optimum cutting-temperature medium until sectioning. A HM 450 Sliding Microtome (Thermo Scientific) was used to section the brains to obtain 30–50- μ m coronal slices. Slices were blocked with 1% bovine serum albumin, 2% normal goat serum and 0.3% TritonX-100 in PBS at room temperature for 1 h and incubated with primary antibodies in working buffer (0.1% bovine serum albumin, 0.2% normal goat serum, 0.3% TritonX-100 in PBS) at 4 °C overnight. The following primary antibodies were used: rabbit anti-RFP (1:200, Abcam), rat anti-RFP (1:300, Chromotek), chicken anti-GFP (1:500, Aves Labs), rabbit anti-GFP (1:2000, Life Technologies) and mouse anti-NeuN (1:200, Millipore). The slices were washed four times with working buffer for 10 min each, incubated with secondary antibodies conjugated with Alexa Fluor 488, 594 or 647 (1:500 or 1:1,000, Life Technologies) in working buffer for 1 h at room temperature, and then washed four times with working buffer for 10 min each. NeuroTrace 435/455 blue fluorescent Nissl stain (1:200, Life Technologies) was used to label neurons after antibody staining. Slices were mounted in Vectashield Mounting Medium containing 4',6-diamidino-2-phenylindole (DAPI) (Vector Laboratories) or ProLong Gold antifade reagent (Life Technologies). Images were acquired on an Olympus FV1000 Confocal, a Zeiss

Axio Imager A1 or an Olympus MVX10 Macroview, and processed using National Institutes of Health ImageJ. To estimate the fraction of layer 2/3 pyramidal cells that were transfected by *in utero* electroporation (Fig. 3a), transfected neurons (tdTomato+) and total neurons (NeuN+) in layer 2/3 were visually quantified. Assuming that 13.2% of layer 2/3 neurons are inhibitory interneurons (Extended Data Fig. 3), we estimated that $9 \pm 1\%$ (mean \pm s.e.m., $n = 12$ sections from six mice) of layer 2/3 pyramidal cells were transfected.

In vitro physiology. Mice between postnatal days 14 and 23 were anaesthetized by an intraperitoneal injection of a ketamine and xylazine mix (100 mg kg^{-1} and 10 mg kg^{-1} , respectively), and transcardially perfused with cold ($0-4^\circ\text{C}$) slice cutting solution containing 80 mM NaCl, 2.5 mM KCl, 1.3 mM NaH_2PO_4 , 26 mM NaHCO_3 , 4 mM MgCl_2 , 0.5 mM CaCl_2 , 20 mM D-glucose, 75 mM sucrose and 0.5 mM sodium ascorbate (315 mosmol, pH 7.4, saturated with 95% $\text{O}_2/5\%$ CO_2). Brains were removed and sectioned in the cutting solution with a Super Microslicer Zero1 (D.S.K.) to obtain 300- μm coronal slices. Slices were incubated in a custom-made interface holding chamber saturated with 95% $\text{O}_2/5\%$ CO_2 at 34°C for 30 min and then at room temperature for 20 min to 8 h until they were transferred to the recording chamber.

Recordings were performed on submerged slices in artificial cerebrospinal fluid (ACSF) containing 119 mM NaCl, 2.5 mM KCl, 1.3 mM NaH_2PO_4 , 26 mM NaHCO_3 , 1.3 mM MgCl_2 , 2.5 mM CaCl_2 , 20 mM D-glucose and 0.5 mM sodium ascorbate (300 mosmol, pH 7.4, saturated with 95% $\text{O}_2/5\%$ CO_2 , perfused at 3 ml min^{-1}) at $31-32^\circ\text{C}$. For whole-cell recordings, we used a K^+ -based pipette solution containing 142 mM K^+ -gluconate, 10 mM HEPES, 1 mM EGTA, 2.5 mM MgCl_2 , 4 mM ATP-Mg, 0.3 mM GTP-Na, 10 mM Na_2 -phosphocreatine (295 mosmol, pH 7.35) or a Cs^+ -based pipette solution containing 115 mM Cs^+ -methanesulphonate, 10 mM HEPES, 1 mM EGTA, 1.5 mM MgCl_2 , 4 mM ATP-Mg, 0.3 mM GTP-Na, 10 mM Na_2 -phosphocreatine, 2 mM QX 314-Cl, 10 mM BAPTA-tetracesium (295 mosmol, pH 7.35). Membrane potentials were not corrected for liquid junction potential (experimentally measured as 11.4 mV for the K^+ -based pipette solution and 8.4 mV for the Cs^+ -based pipette solution).

Neurons were visualized with video-assisted infrared differential interference contrast imaging and fluorescent neurons were identified by epifluorescence imaging under a water immersion objective ($\times 40$, 0.8 numerical aperture) on an upright Olympus BX51WI microscope with an infrared CCD camera (VX44, Till Photonics). For Fos-EGFP experiments, in a given field-of-view those pyramidal cells with the strongest EGFP fluorescence were visually identified as the EGFP+ neurons. The EGFP- neurons were those pyramidal cells whose fluorescence was equal to the background fluorescence level of the slices.

Data were low-pass filtered at 4 kHz and acquired at 10 kHz with an Axon Multiclamp 700A or 700B amplifier and an Axon Digidata 1440A Data Acquisition System under the control of Clampex 10.2 (Molecular Devices). Data were analysed offline using AxoGraph X (AxoGraph Scientific).

For the photostimulation of ChR2-expressing neurons, blue light was emitted from a collimated light-emitting diode (470 nm) driven by a T-Cube LED Driver (Thorlabs) under the control of an Axon Digidata 1440A Data Acquisition System and Clampex 10.2. Light was delivered through the reflected light fluorescence illuminator port and the $\times 40$ objective.

Synaptic currents were recorded in the whole-cell voltage clamp mode with the Cs^+ -based patch pipette solution. Only recordings with series resistance below 20 M Ω were included. EPSCs and IPSCs were recorded at the reversal potential for IPSCs (-60 mV) and EPSCs ($+10 \text{ mV}$), respectively. For light pulse stimulation, pulse duration (0.5–5 ms) and intensity ($1.1-5.5 \text{ mW mm}^{-2}$) were adjusted for each recording to evoke small (to minimize voltage-clamp errors; see the figures for the ranges) but reliable monosynaptic EPSCs or IPSCs. Disynaptic IPSCs were evoked using the same light pulses that were used for evoking the corresponding monosynaptic EPSCs. Light pulses were delivered at 30-s interstimulus intervals.

To quantify the inter-cell variability of EPSCs (Fig. 1d, f), we used the average relative deviation defined as $\frac{1}{N \times \text{EPSC}_{\text{mean}}} \sum_{i=1}^N |\text{EPSC}_i - \text{EPSC}_{\text{mean}}|$, where N is the number of pyramidal cells in one given experiment, EPSC_i is the amplitude of the EPSC recorded in the i^{th} pyramidal cell within that experiment and $\text{EPSC}_{\text{mean}}$ is the mean amplitude of EPSCs recorded across pyramidal cells in the same experiment. The average relative deviation of IPSCs or E/I ratios was obtained in the same way for each experiment (Fig. 1d, f).

To record unitary connections between inhibitory neurons and pyramidal cells, Pvalb and Sst cells were identified by the Cre-dependent expression of ChR2-tdTomato or hChR2(H134R)-EYFP in *Pvalb-ires-Cre* and *Sst-ires-Cre* mice, respectively. Pyramidal cells were first recorded in whole-cell voltage clamp mode ($+10 \text{ mV}$) with the Cs^+ -based patch pipette solution, and a nearby Pvalb or Sst cell was subsequently recorded in the whole-cell current clamp mode with the K^+ -based patch pipette solution. Action potentials were elicited in Pvalb or Sst cells by a 2-ms depolarizing current step (1–2 nA) with a 15-s interstimulus intervals.

Unitary IPSC (uIPSC) amplitudes were measured from the average of 10–50 sweeps. We considered a Pvalb or Sst cell to be connected with a pyramidal cell when the average uIPSC amplitude was at least three times the baseline standard deviation. The average relative deviation of uIPSC amplitudes (Fig. 5 and Extended Data Fig. 9) was calculated as $\frac{1}{N \times \text{uIPSC}_{\text{mean}}} \sum_{i=1}^N |\text{uIPSC}_i - \text{uIPSC}_{\text{mean}}|$ where N is

the number of pyramidal cells in one given experiment, uIPSC_i is the amplitude of the IPSC recorded in the i^{th} pyramidal cell within that experiment and $\text{uIPSC}_{\text{mean}}$ is the mean amplitude of uIPSCs recorded across pyramidal cells in the same experiment.

Neuronal intrinsic excitability was examined with the K^+ -based pipette solution in the presence of the AMPA receptor antagonist NBQX (10 μM), the NMDA receptor antagonist (RS)-CPP (10 μM) and the GABA_A (γ -aminobutyric acid) receptor antagonist SR 95531 (10 μM). The resting membrane potential was recorded in the whole-cell current clamp mode within the first minute after break-in. The input resistance was measured after balancing the bridge by injecting a 500-ms-long hyperpolarizing current pulse (10–100 pA) to generate a small membrane potential hyperpolarization (2–10 mV) from the resting membrane potential. Depolarizing currents were increased in 5- or 10-pA steps to identify rheobase currents.

Ba^{2+} -sensitive currents were measured with the K^+ -based pipette solution in the presence of NBQX (10 μM), (RS)-CPP (10 μM), SR 95531 (10 μM) and Na^+ channel blocker TTX (1 μM). Only recordings with series resistance below 20 M Ω were included. Neurons were clamped at -25 mV and the membrane potential was ramped to -125 mV at a rate of 20 mV s^{-1} . The membrane currents recorded in the presence of BaCl_2 (50 μM) were subtracted from those recorded in the absence of BaCl_2 to obtain the Ba^{2+} -sensitive currents, which were then divided by the whole-cell membrane capacitances to calculate the current densities.

In vivo physiology. Mice between postnatal days 17 and 23 were anaesthetized by an intraperitoneal injection of chlorprothixene (5 mg kg^{-1}) followed by urethane (1.2 g kg^{-1}). Oxygen was given at a flow rate of 1 l min^{-1} during the experiments and isoflurane ($<0.5\%$) was supplemented if necessary. The body temperature was maintained by a feedback-based DC temperature control system at 37°C . Dexamethasone sodium phosphate (2 mg kg^{-1}) and Lactated Ringer's Injection (3 ml kg^{-1} every 2 h) were administered subcutaneously. Whiskers and eyelashes were trimmed, and a thin layer of silicone oil (kinematic viscosity 30,000 centistokes ($1 \text{ cSt} = 10^{-6} \text{ m}^2 \text{ s}^{-1}$), Sigma-Aldrich) was applied to the corneas to prevent drying. The scalp and periosteum were removed. Vetbond tissue adhesive (3M) was applied to stabilize all sutures. A custom-made recording chamber with a 3-mm diameter hole in the centre was attached to the skull over V1 with Vetbond tissue adhesive and dental cement (Ortho-Jet BCA, Lang Dental). The recording chamber was then secured on a custom-made holder. A craniotomy (1.5–2 mm diameter, centred at 2.5 mm lateral to midline and 1 mm rostral to lambda suture) was performed with a 0.3-mm diameter round bur on a high-speed rotary micromotor. The dura was left intact and the craniotomy was covered by a thin layer of 1.5% type III-A agarose in HEPES-ACSF containing 142 mM NaCl, 5 mM KCl, 10 HEPES-Na, 1.3 mM MgCl_2 , 3.1 mM CaCl_2 and 10 mM D-glucose (310 mosmol, pH 7.4). HEPES-ACSF was added to the recording chamber.

Targeted loose-patch recordings were performed under the guidance of a two-photon laser scanning microscope. Two-photon imaging was performed with a water immersion objective ($\times 40$, 0.8 numerical aperture, Olympus) on a Moveable Objective Microscope (Sutter Instruments) coupled with a Ti:Sapphire laser (Chameleon Ultra II, Coherent) under the control of ScanImage 3.6 (Janelia Farm Research Campus, HHMI)⁴⁷. Laser wavelength was tuned to 950 nm (laser power after the objective: 25–50 mW) for two-photon excitation of tdTomato and Alexa Fluor 488.

An Axon Multiclamp 700B amplifier was used for extracellular recording of spikes. A patch pipette containing HEPES-ACSF and 10–20 μM Alexa Fluor 488 hydrazide (Life Technologies) was advanced along its axis towards neurons located between 150 and 250 μm below the dura with minimal lateral movements. A small positive pressure was applied to the patch pipette to avoid clogging of the tip and to inject a small amount of fluorescent dye to stain the extracellular space. Non-fluorescent neurons were visualized as negative images⁴⁸. The pipette resistance was constantly monitored in voltage-clamp mode. The concurrence of the pipette tip contacting the neuron and an increase in pipette resistance indicated successful targeting, which was further confirmed post hoc (see below). Upon the release of positive pressure, a small negative pressure was applied to form a loose seal (10–30 M Ω). The amplifier was then switched to the current-clamp mode with zero current injection to record voltage. Data were low-pass filtered at 10 kHz and acquired at 32 kHz with a NI-DAQ board (NI PCIe-6259, National Instruments) under the control of a custom-written program running in Matlab (Mathworks). Within a local region ($<50 \mu\text{m}$), neighbouring tdTomato⁺ and tdTomato[−] neurons were sequentially targeted for recording, but the order of recordings from tdTomato⁺ and tdTomato[−] was alternated. The correct targeting of tdTomato⁺ neurons was confirmed at the end of the recording either by the filling of the neuron with the

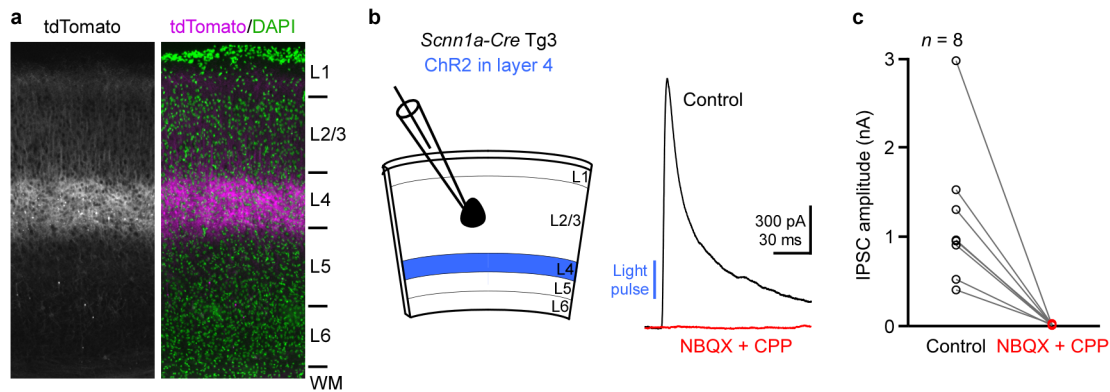
fluorescent dye contained in the pipette via break-in or by the presence of neuronal fluorescence in the recording pipette due to the negative pressure.

Visual stimuli were generated in Matlab with Psychophysics Toolbox⁴⁹ and displayed on a gamma-corrected liquid-crystal display monitor (30 cm × 47.5 cm, 60 Hz refresh rate, mean luminance 50 cd m⁻²). The monitor was placed 25 cm away from the contralateral eye, covering 62° (vertical) × 87° (horizontal) of the visual space. The monitor was approximately centred at the retinotopic location corresponding to the V1 recording site by monitoring single-unit or multi-unit activity in response to a moving bar on the screen. During recordings, full-field sinusoidal drifting gratings (temporal frequency 2 Hz, spatial frequency 0.04 cycles per degree, 100% contrast) were presented randomly at 12 different directions from 0° to 330° for 1.5 s, preceded and followed by the presentation of a grey screen for 2 s and 1.5 s, respectively. The complete set of stimuli was repeated 8–16 times.

We analysed data offline using a custom-written program in Matlab. Voltage signals were high-pass filtered (125 Hz). Spikes were first detected as events exceeding five times the standard deviation of the noise, and then visually verified. Spontaneous spike rate was calculated as the average spike rate during the 2-s time window before the presentation of a visual stimulus. Evoked spike rate was calculated as the average spike rate during the 1.5-s time window of visual stimulation. Overall spike rate was calculated as the average spike rate during the entire recording period.

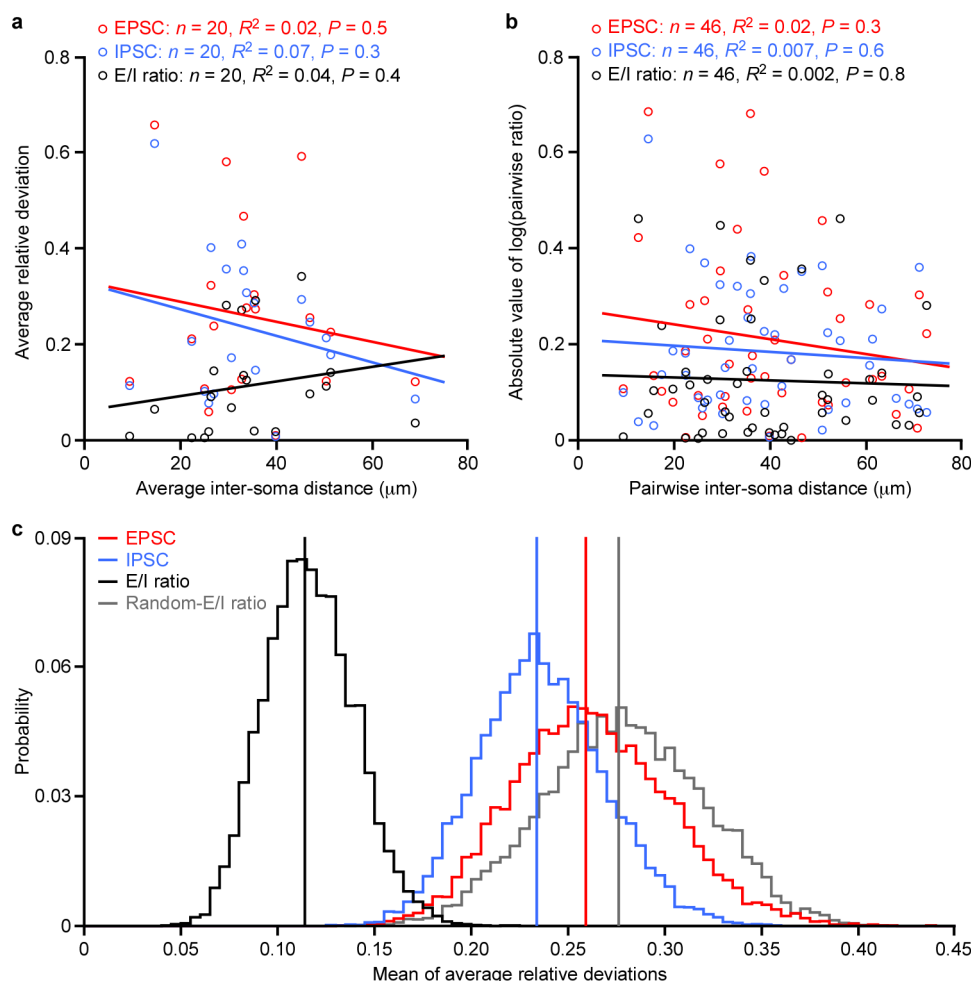
Statistics. All reported sample numbers (*n*) represent biological replicates. Sample sizes were estimated to have 70–80% power to detect expected effect size using StatMate 2 (GraphPad Software). Statistical analyses used Prism 5 (GraphPad Software) and Matlab. Linear regression with an *F*-test (two-sided) was used for Fig. 1e and Extended Data Fig. 2a, b. Bootstrapped distributions (Extended Data Fig. 2c) were used to determine the statistical significance for Fig. 1f. A Wilcoxon matched-pairs signed rank test (two-sided) was used for Figs 2, 3h–j, 4, 5b and Extended Data Figs 1–5. A Mann–Whitney *U*-test (two-sided) was used for Fig. 3d–f, uIPSC amplitudes in Figs 5f, g and Extended Data Figs 4, 6, 9c, d. Fisher's exact test (two-sided) was used for connectivity rates in Fig. 5f and Extended Data Fig. 9b.

31. Madisen, L. *et al.* A robust and high-throughput Cre reporting and characterization system for the whole mouse brain. *Nature Neurosci.* **13**, 133–140 (2010).
32. Barth, A. L., Gerkin, R. C. & Dean, K. L. Alteration of neuronal firing properties after *in vivo* experience in a FosGFP transgenic mouse. *J. Neurosci.* **24**, 6466–6475 (2004).
33. Taniguchi, H. *et al.* A resource of Cre driver lines for genetic targeting of GABAergic neurons in cerebral cortex. *Neuron* **71**, 995–1013 (2011).
34. Hippenmeyer, S. *et al.* A developmental switch in the response of DRG neurons to ETS transcription factor signaling. *PLoS Biol.* **3**, e159 (2005).
35. Yang, J., Jan, Y. N. & Jan, L. Y. Control of rectification and permeation by residues in two distinct domains in an inward rectifier K⁺ channel. *Neuron* **14**, 1047–1054 (1995).
36. Tong, Y. *et al.* Tyrosine decaging leads to substantial membrane trafficking during modulation of an inward rectifier potassium channel. *J. Gen. Physiol.* **117**, 103–118 (2001).
37. Tinker, A., Jan, Y. N. & Jan, L. Y. Regions responsible for the assembly of inwardly rectifying potassium channels. *Cell* **87**, 857–868 (1996).
38. Ren, D. *et al.* A prokaryotic voltage-gated sodium channel. *Science* **294**, 2372–2375 (2001).
39. Yue, L., Navarro, B., Ren, D., Ramos, A. & Clapham, D. E. The cation selectivity filter of the bacterial sodium channel, NaChBac. *J. Gen. Physiol.* **120**, 845–853 (2002).
40. Turan, S., Kuehle, J., Schambach, A., Baum, C. & Bode, J. Multiplexing RMCE: versatile extensions of the Flp-recombinase-mediated cassette-exchange technology. *J. Mol. Biol.* **402**, 52–69 (2010).
41. Kranz, A. *et al.* An improved Flp deleter mouse in C57Bl/6 based on Flpo recombinase. *Genesis* **48**, 512–520 (2010).
42. Matsuda, T. & Cepko, C. L. Electroporation and RNA interference in the rodent retina *in vivo* and *in vitro*. *Proc. Natl Acad. Sci. USA* **101**, 16–22 (2004).
43. Manent, J. B., Wang, Y., Chang, Y., Paramasivam, M. & LoTurco, J. J. Dcx reexpression reduces subcortical band heterotopia and seizure threshold in an animal model of neuronal migration disorder. *Nature Med.* **15**, 84–90 (2009).
44. Matsuda, T. & Cepko, C. L. Controlled expression of transgenes introduced by *in vivo* electroporation. *Proc. Natl Acad. Sci. USA* **104**, 1027–1032 (2007).
45. Atasoy, D., Aponte, Y., Su, H. H. & Sternson, S. M. A FLEX switch targets channelrhodopsin-2 to multiple cell types for imaging and long-range circuit mapping. *J. Neurosci.* **28**, 7025–7030 (2008).
46. Sohal, V. S., Zhang, F., Yizhar, O. & Deisseroth, K. Parvalbumin neurons and gamma rhythms enhance cortical circuit performance. *Nature* **459**, 698–702 (2009).
47. Pologruto, T. A., Sabatini, B. L. & Svoboda, K. ScanImage: flexible software for operating laser scanning microscopes. *Biomed. Eng. Online* **2**, 13 (2003).
48. Kitamura, K., Judkewitz, B., Kano, M., Denk, W. & Hausser, M. Targeted patch-clamp recordings and single-cell electroporation of unlabeled neurons *in vivo*. *Nature Methods* **5**, 61–67 (2008).
49. Brainard, D. H. The Psychophysics Toolbox. *Spat. Vis.* **10**, 433–436 (1997).
50. Hibino, H. *et al.* Inwardly rectifying potassium channels: their structure, function, and physiological roles. *Physiol. Rev.* **90**, 291–366 (2010).



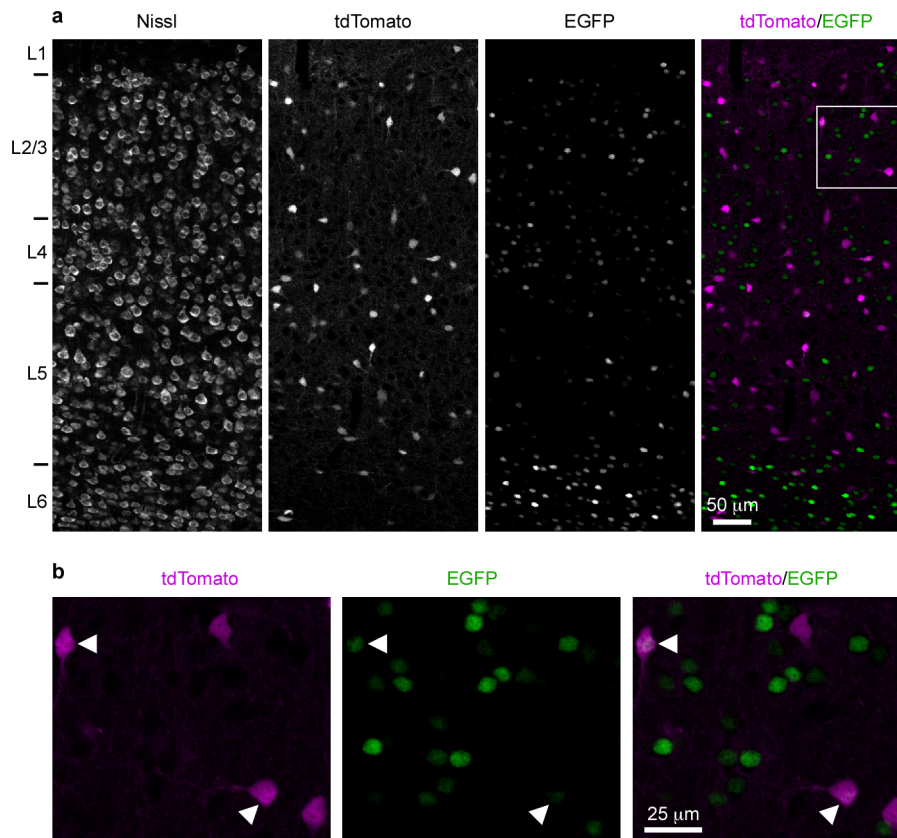
Extended Data Figure 1 | Cre recombinase-expressing cells in the cortex of *Scnn1a-Cre-Tg3* mice are layer 4 excitatory neurons. AAV-CAGGS-Flex-ChR2-tdTomato, expressing ChR2-tdTomato fusion protein in a Cre-dependent manner, was injected into *Scnn1a-Cre-Tg3* mice. **a**, Representative fluorescent images of a coronal section of V1 showing that the ChR2-tdTomato-expressing cells located primarily in layer 4 ($n = 11$ mice). Cortical layers are indicated on the right based on the DAPI staining pattern. L, layer; WM, white matter. **b**, Left, schematic of experiments. Right, a layer 2/3

pyramidal cell was voltage clamped at the reversal potential for excitation (+10 mV). Photoactivation of ChR2-expressing neurons in layer 4 elicited an IPSC (black trace), which was abolished by the glutamatergic receptor antagonists NBQX and CPP (red trace), indicating its disynaptic nature. **c**, Summary data: NBQX and CPP reduced IPSC amplitudes by $98.0 \pm 0.6\%$ (mean \pm s.e.m., $n = 8$, $P = 0.008$) indicating that ChR2 was exclusively expressed in excitatory neurons.



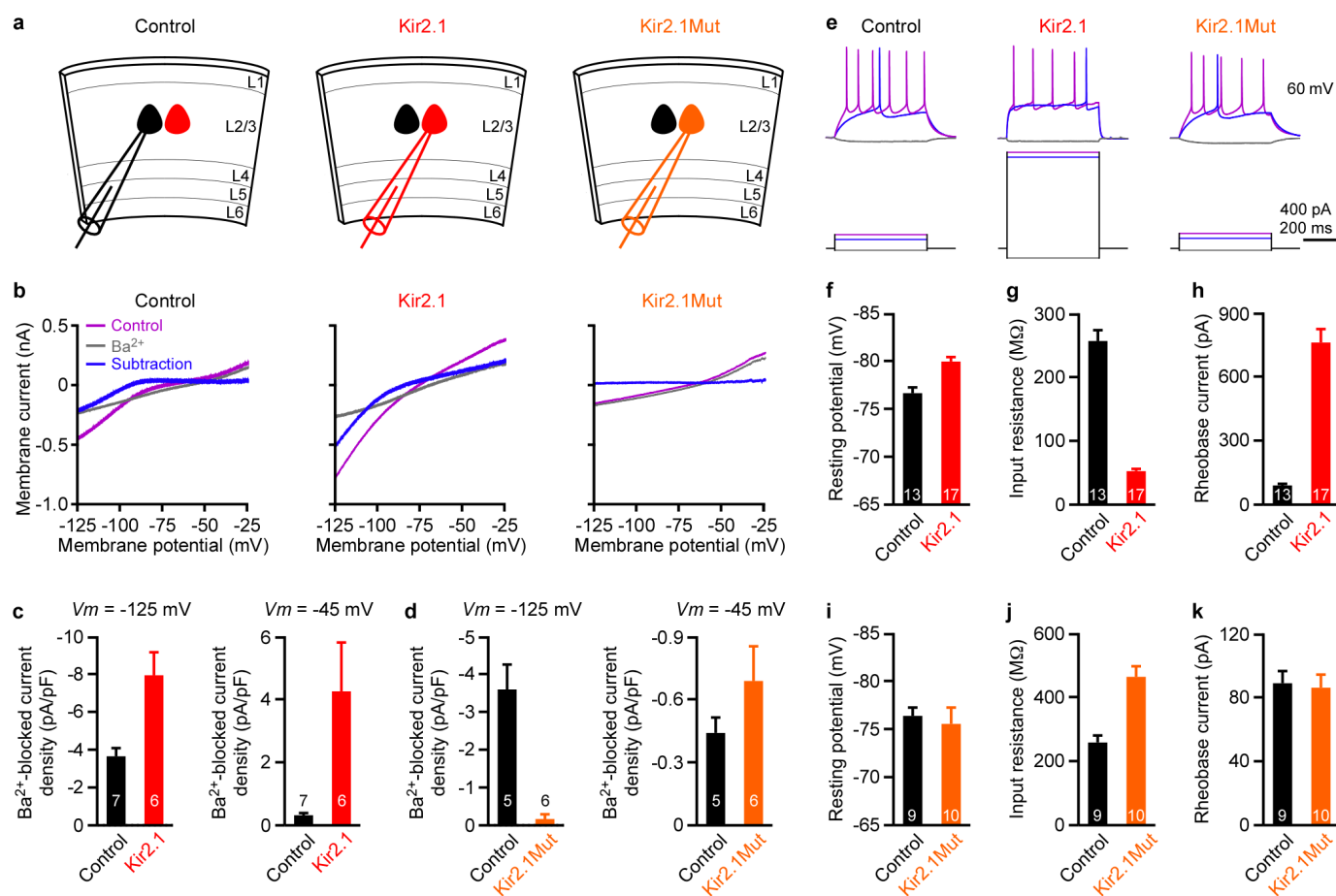
Extended Data Figure 2 | Characterization of the inter-cell variability of EPSCs, IPSCs and E/I ratios. **a, b,** The inter-cell variability of EPSCs, IPSCs and E/I ratios among neighbouring pyramidal cells does not correlate with their inter-soma distances. **a,** The average relative deviations of EPSCs, IPSCs and E/I ratios from each experiment in Fig. 1f are plotted against the average inter-soma distance from the same experiment. The average inter-soma distance is the mean of the distances between each pair of pyramidal cells. For the experiments in which only two pyramidal cells were recorded, the inter-soma distance between the two pyramidal cells was used. Lines, linear regression fits. **b,** The absolute value of the logarithm of the ratio of EPSCs (or IPSCs or E/I ratios) simultaneously recorded in two pyramidal cells was plotted against the inter-somatic distance between the two cells. **c,** The distribution of E/I ratios across pyramidal cells varies less than if EPSCs and IPSCs were randomly paired between cells and less than the distributions of EPSC and IPSC amplitudes. To determine whether the precise E/I ratio recorded within each pyramidal cell minimizes the average relative deviation, we computed the E/I ratios from randomly but uniquely paired EPSCs and IPSCs within each of the 20 experiments from Fig. 1f. By randomizing within each experiment, we ensured that the average relative deviation was only modified owing to the pairing of EPSCs to IPSCs. Note that, for an experiment with N pyramidal cells, there were $N!$ possible randomized pairings of EPSCs and IPSCs, and hence $N!$ possible E/I ratio average relative deviations (referred to as random-E/I ratio

average relative deviations). The distribution of the means of the random-E/I ratio average relative deviations (grey histogram) was constructed from the means of 10,000 samples. Each sample consisted of 20 random-E/I ratio average relative deviations, each of which was randomly chosen from the $N!$ possible random-E/I ratio average relative deviations of each experiment. The grey vertical line represents the mean of the distribution. The distribution of the means of the E/I ratio average relative deviations (black histogram) was generated by bootstrapping (that is, resampling 10,000 times with replacement). Each resample consisted of 20 randomly chosen E/I ratio average relative deviations from the 20 experiments in Fig. 1f, and an E/I ratio average relative deviation was allowed to be repeated within one resample (that is, sampling with replacement). The black vertical line represents the mean of the experimentally obtained E/I ratio average relative deviations. The E/I ratio average relative deviations are smaller than the random-E/I ratio average relative deviations ($P < 0.0001$). The distributions of the means of the EPSC average relative deviations (red histogram) and the means of the IPSC average relative deviations (blue histogram) were generated by similar bootstrapping to the E/I ratio average relative deviations. The red and blue vertical lines represent the means of the experimentally obtained EPSC average relative deviations and IPSC average relative deviations, respectively. The E/I ratio average relative deviations are smaller than the EPSC average relative deviations ($P < 0.0001$) and the IPSC average relative deviations ($P < 0.0001$).



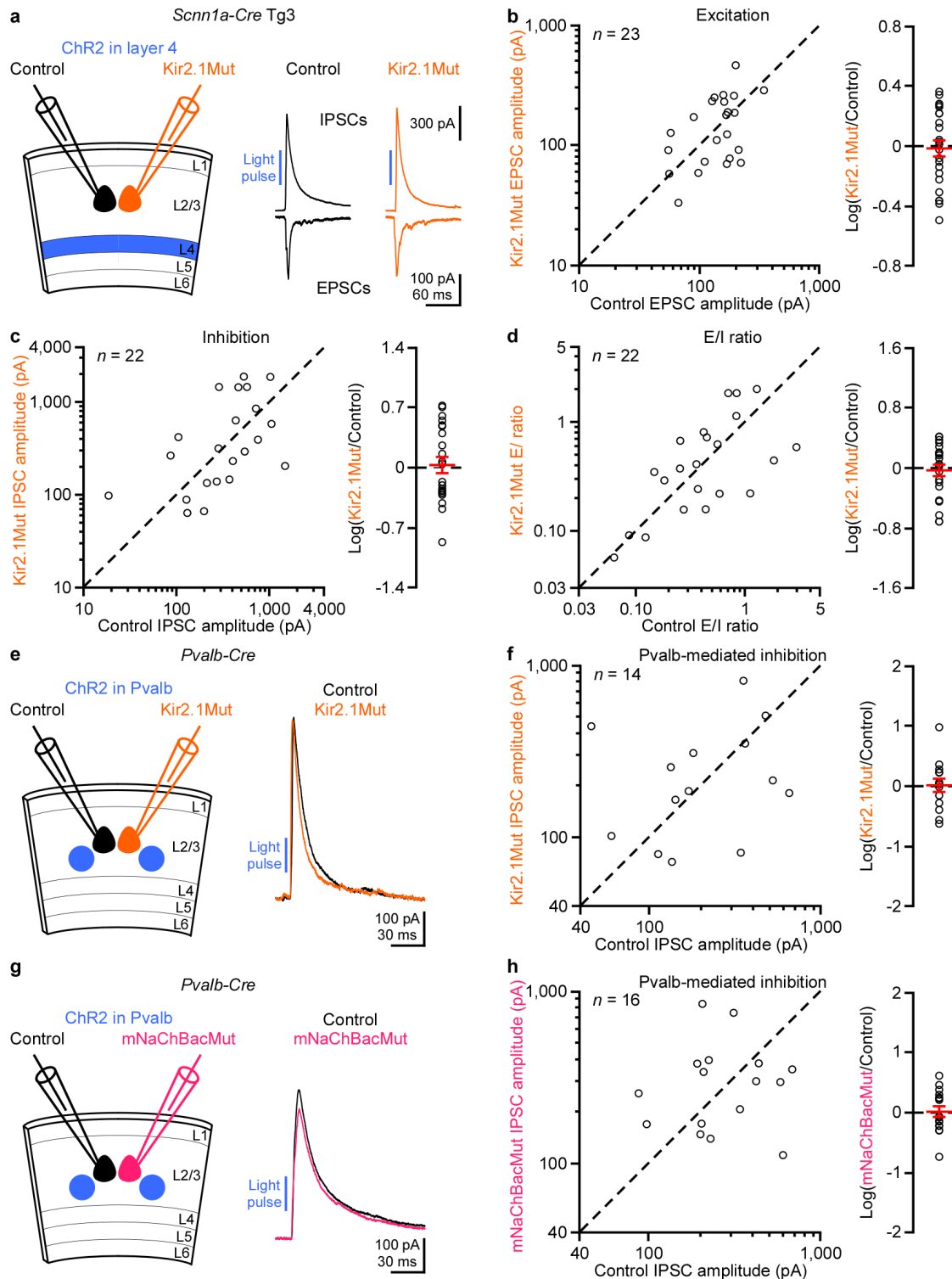
Extended Data Figure 3 | Most layer 2/3 Fos-EGFP+ neurons in V1 are pyramidal cells. *Fos-EGFP* mice were crossed with *Gad2-ires-Cre* and *Rosa-CAG-LSL-tdTomato-WPRE* mice to generate *Fos-EGFP, Gad2-ires-Cre, Rosa-CAG-LSL-tdTomato-WPRE* mice. **a**, Representative fluorescent images showed a coronal section of V1. All neurons were visualized by NeuroTrace 435/455 blue fluorescent Nissl stain and GABAergic interneurons were labelled by tdTomato. EGFP was stained with an antibody against GFP and visualized with a secondary antibody conjugated with Alexa Fluor 647. Cortical layers are

indicated on the left based on the Nissl staining pattern. **b**, Enlarged view of the boxed region in **a**. In layer 2/3 of V1, only $5.3 \pm 0.9\%$ (mean \pm s.e.m., $n = 10$ sections from two mice) of EGFP+ neurons were GABAergic interneurons (two examples are indicated by arrowheads). GABAergic interneurons constitute $13.2 \pm 0.6\%$ (mean \pm s.e.m., $n = 14$ sections from three mice including one *Gad2-ires-Cre, Rosa-CAG-LSL-tdTomato-WPRE* mouse) of all layer 2/3 neurons.



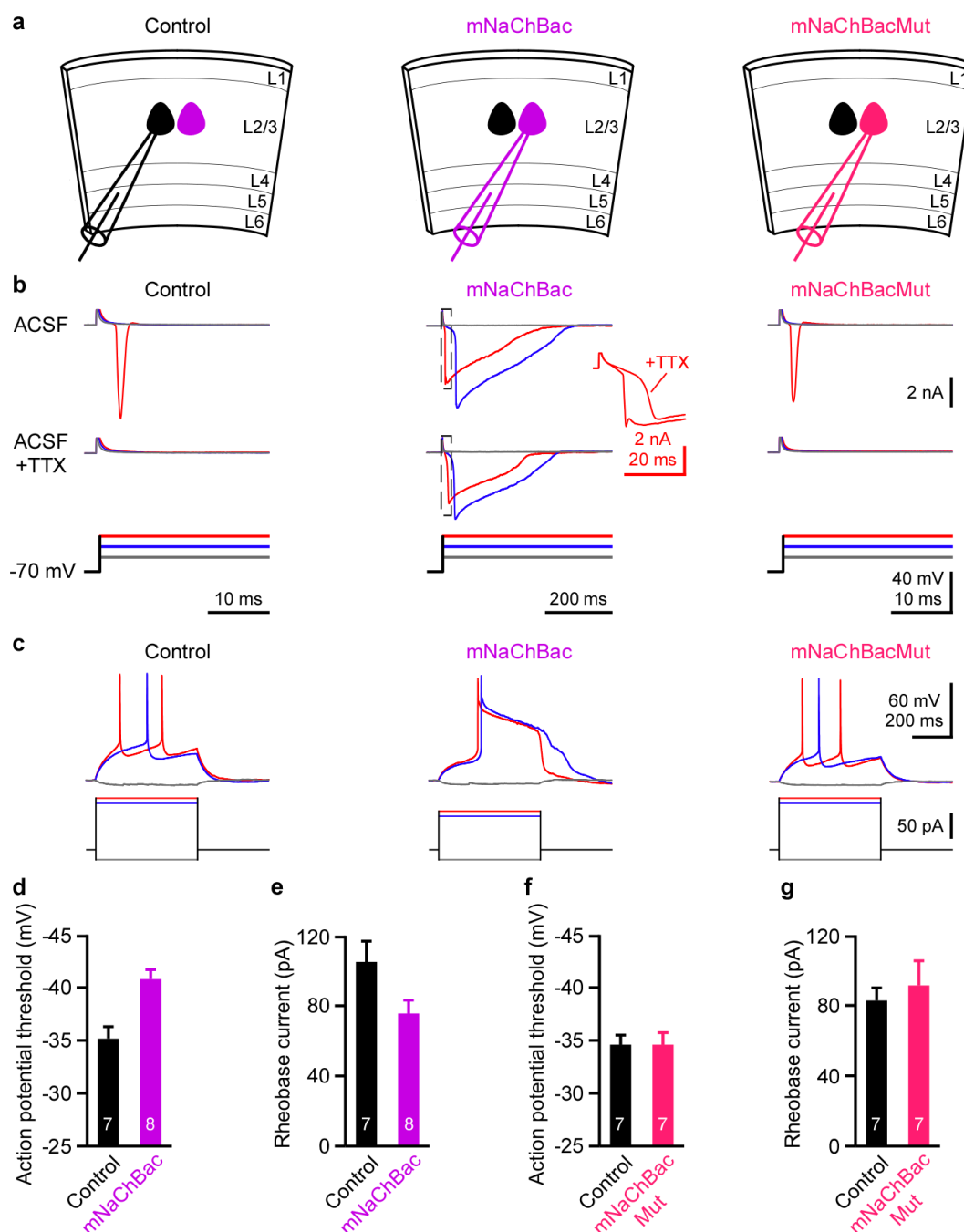
Extended Data Figure 4 | Overexpression of Kir2.1 increases a Ba^{2+} -sensitive K^+ current and decreases neuronal excitability. **a**, Schematics of experiments. Kir2.1 or a non-conducting mutant Kir2.1 (Kir2.1Mut) was overexpressed in a subset of layer 2/3 pyramidal cells by *in utero* electroporation. **b**, Membrane currents in response to a 5 s membrane potential ramp from -25 to -125 mV from an untransfected control pyramidal cell, a pyramidal cell overexpressing Kir2.1 and a pyramidal cell overexpressing Kir2.1Mut. The purple traces were recorded in control condition and the grey traces were recorded in the presence of 50 μM BaCl_2 , a concentration that primarily blocks the K^+ channels of the Kir2 subfamily⁵⁰. The blue traces were obtained by subtracting the grey traces from the purple traces, representing the Ba^{2+} -blocked K^+ currents. **c**, The exogenously overexpressed Kir2.1 increased not only the Ba^{2+} -blocked inward current density at -125 mV ($P = 0.01$), but also the outward current density at -45 mV ($P = 0.001$) owing

to its reduced inward rectification (see Methods). **d**, Kir2.1Mut can bind to the endogenous Kir2.1 to form non-conducting channels⁵⁰, acting as a dominant negative to decrease the inward current density at -125 mV ($P = 0.004$) but without affecting the outward current density at -45 mV ($P = 0.2$). **e**, Membrane potentials (upper panels) in response to current injections (lower panels) from an untransfected control pyramidal cell, a pyramidal cell overexpressing Kir2.1 and a pyramidal cell overexpressing Kir2.1Mut. **f-h**, Overexpression of Kir2.1 hyperpolarized the resting membrane potential (**f**, $P = 0.0003$), decreased the resting input resistance (**g**, $P < 0.0001$) and increased the rheobase current (**h**, $P < 0.0001$). **i-k**, Overexpression of Kir2.1Mut increased the resting input resistance (**j**, $P = 0.0002$), but had no effects on the resting membrane potential (**i**, $P = 0.5$) and the rheobase current (**k**, $P = 0.9$). The numbers of recorded neurons are indicated on the bars. All data are expressed as mean \pm s.e.m.



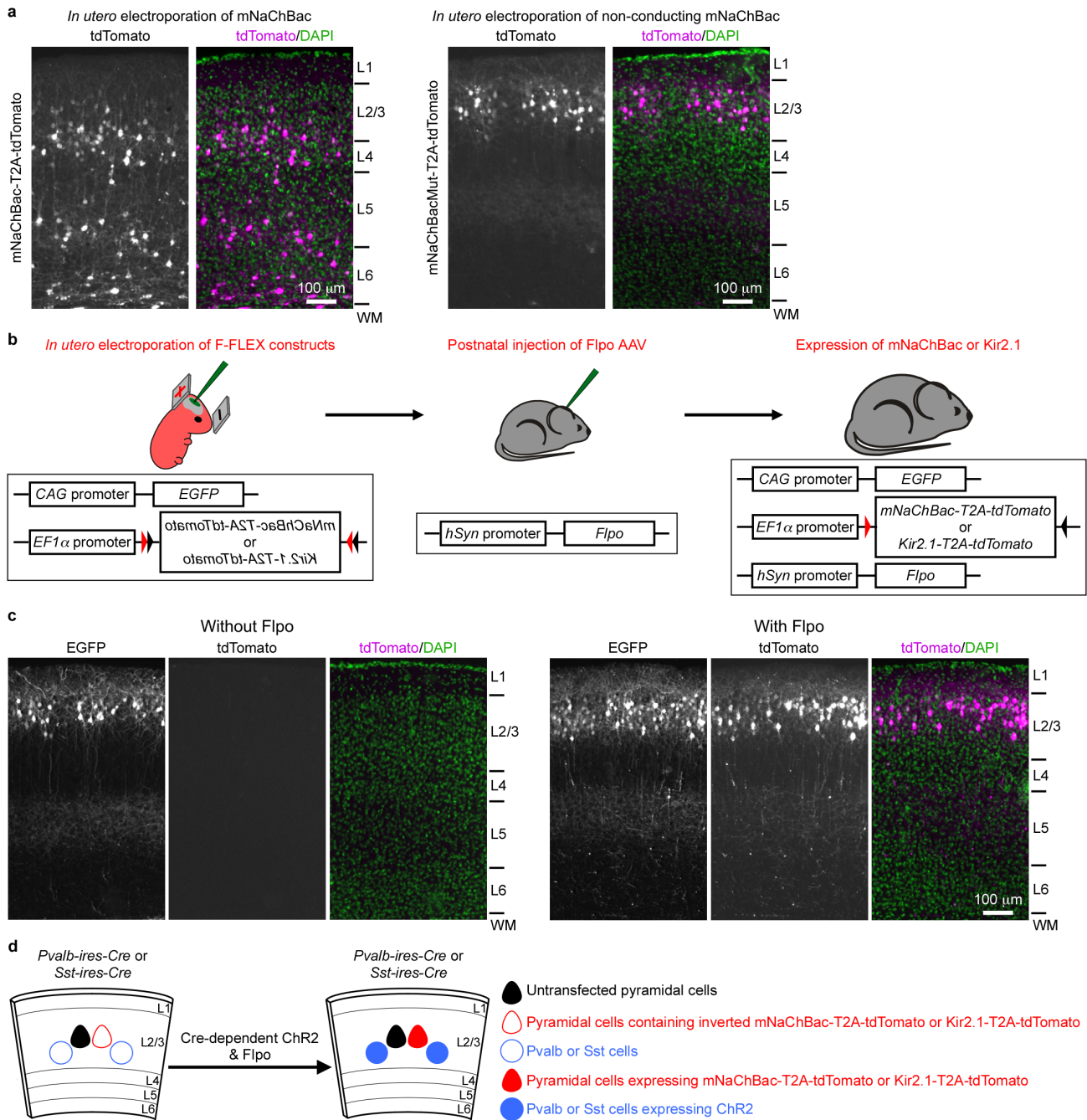
Extended Data Figure 5 | Overexpression of Kir2.1Mut or mNaChBacMut in layer 2/3 pyramidal cells does not affect inhibition. **a**, Left, schematic of experiments. *Scnn1a-Cre-Tg3* mice with ChR2 in layer 4 excitatory neurons and Kir2.1Mut in a subset of layer 2/3 pyramidal cells. Right, monosynaptic EPSCs and disynaptic IPSCs from simultaneously recorded control and Kir2.1Mut neurons in response to layer 4 photoactivation. **b–d**, Summary graphs. **b**, Left, EPSC amplitudes in Kir2.1Mut neurons plotted against those in control neurons. Right, logarithm of the ratio between EPSC amplitudes in Kir2.1Mut and control neurons. Red, mean \pm s.e.m. EPSC amplitudes are similar between Kir2.1Mut and control neurons ($n = 23$, $P = 0.7$). **c**, As in **b**, but for IPSCs. IPSC amplitudes are similar between Kir2.1Mut and control neurons ($n = 22$, $P = 0.6$). **d**, As in **b**, but for E/I ratios. E/I ratios are similar

between Kir2.1 and control neurons ($n = 22$, $P = 0.6$). **e**, Left, schematic of experiments. *Pvalb-ires-Cre* mice with ChR2 in Pvalb cells and Kir2.1Mut in a subset of layer 2/3 pyramidal cells. Right, IPSCs from simultaneously recorded control and Kir2.1Mut neurons in response to Pvalb cell photoactivation. **f**, Summary graphs. Left, IPSC amplitudes in Kir2.1Mut neurons plotted against those in control neurons. Right, logarithm of the ratio between IPSC amplitudes in Kir2.1Mut and control neurons. Red, mean \pm s.e.m. IPSC amplitudes are similar between Kir2.1Mut and control neurons ($n = 14$, $P = 0.8$). **g**, **h**, As in **e**, **f**, but for a non-conducting mutant mNaChBac (mNaChBacMut). IPSC amplitudes are similar between mNaChBacMut and control neurons ($n = 16$, $P = 0.9$).



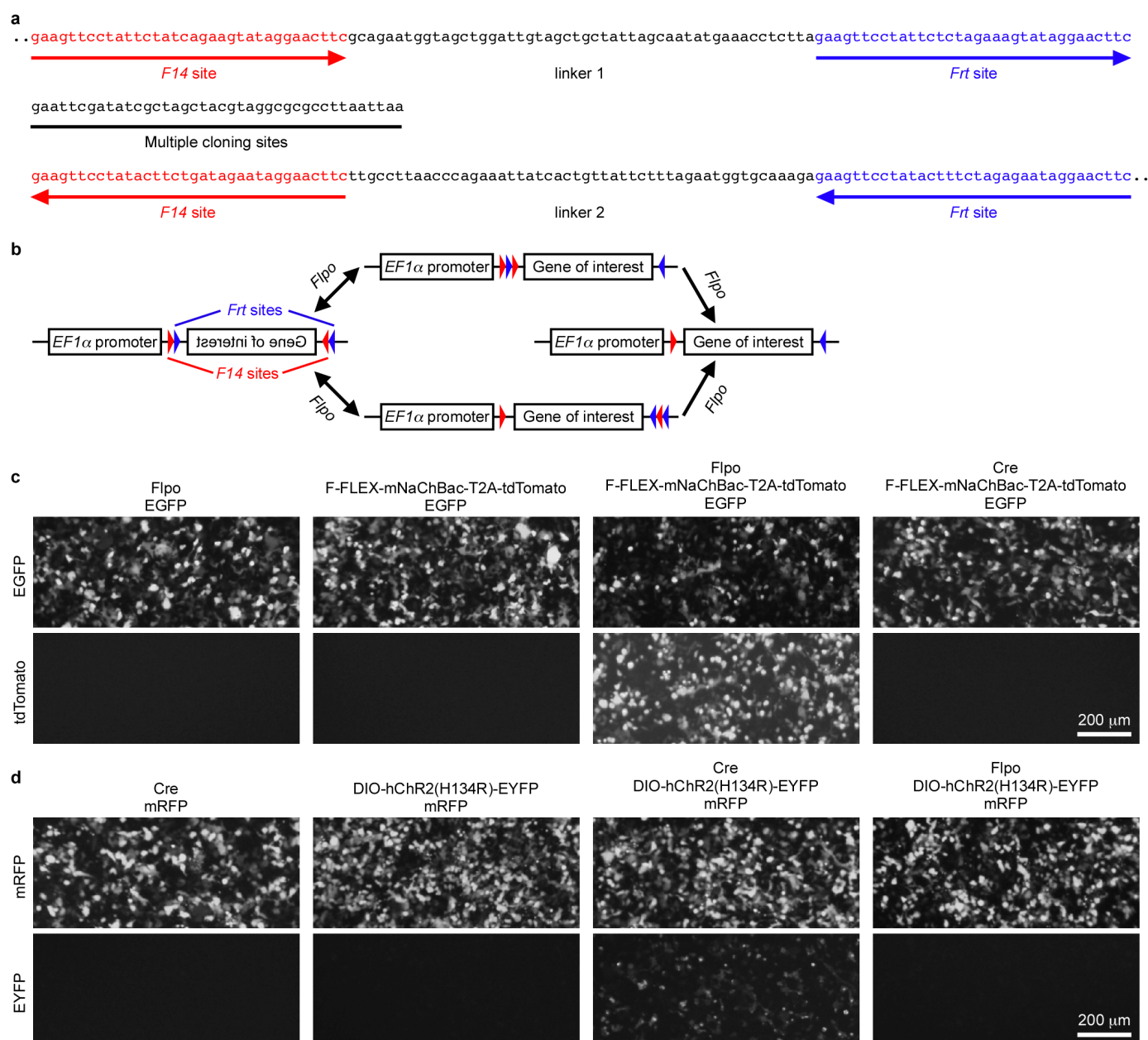
Extended Data Figure 6 | Overexpression of mNaChBac increases neuronal excitability. **a**, Schematics of experiments. mNaChBac or a non-conducting mutant mNaChBac (mNaChBacMut) was overexpressed in a subset of layer 2/3 pyramidal cells by *in utero* electroporation. **b**, Membrane currents (upper and middle panels) in response to voltage steps (lower panels) from an untransfected control pyramidal cell, a pyramidal cell overexpressing mNaChBac and a pyramidal cell overexpressing mNaChBacMut. The endogenous voltage-gated inward Na^+ current was fast inactivating and was blocked by tetrodotoxin (TTX, 1 μM), whereas the mNaChBac-mediated inward current was slow inactivating and insensitive to TTX. Inset, overlay of the two dashed boxes. Note that the fast component of the inward current representing the endogenous Na^+ current was blocked by TTX. **c**, Membrane

potentials (upper panels) in response to current injections (lower panels) from a control neuron, a mNaChBac neuron and a mNaChBacMut neuron. The mNaChBac neuron generated long-lasting action potentials and depolarizations, whereas the mNaChBacMut neuron generated action potentials similar to the control neuron. **d**, **e**, Overexpression of mNaChBac lowered the action potential threshold (defined as the membrane potential whose derivative reaches 2 V s^{-1}) (**d**, $P = 0.004$) and decreased the rheobase current (**e**, $P = 0.03$). **f**, **g**, Overexpression of mNaChBacMut did not alter the action potential threshold (**f**, $P = 0.9$) and the rheobase current (**g**, $P = 0.8$). The numbers of recorded neurons are indicated on the bars. All data are expressed as mean \pm s.e.m.



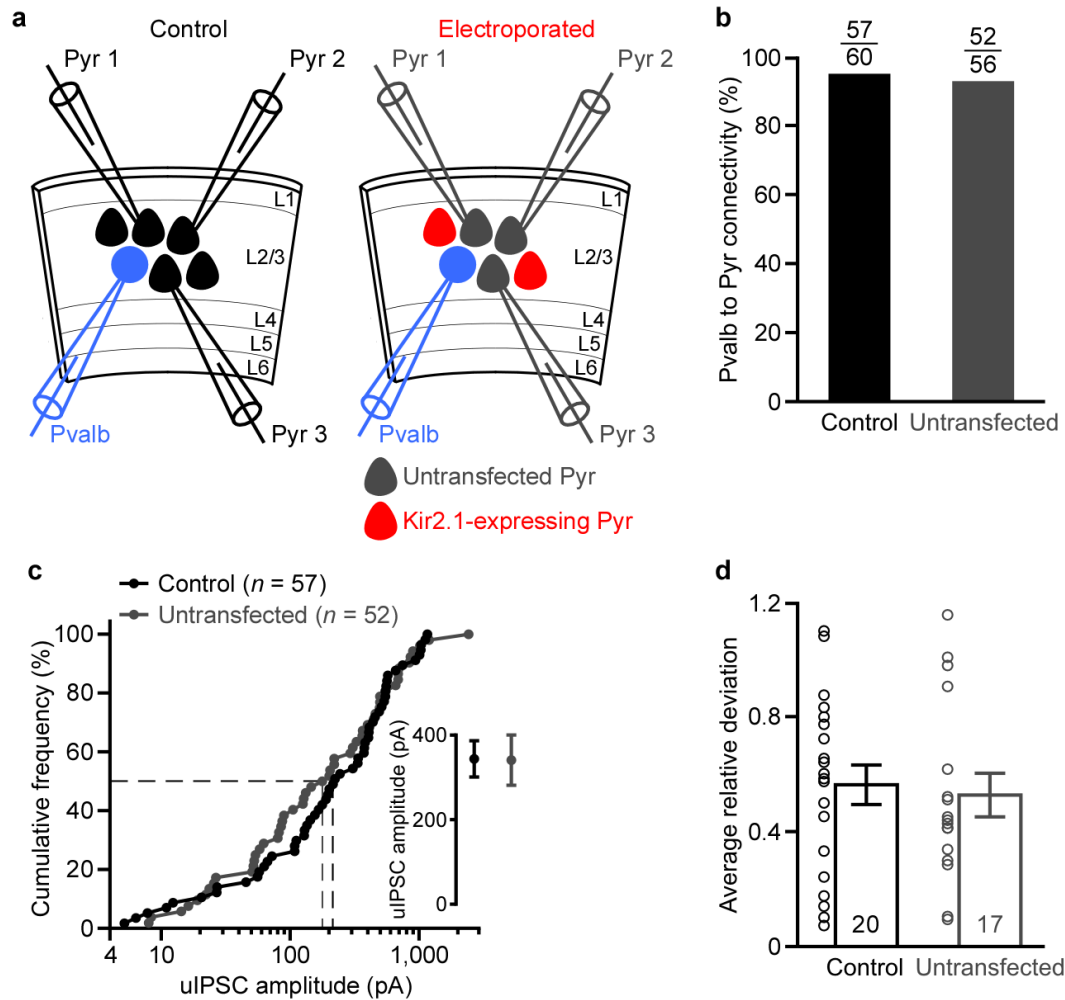
Extended Data Figure 7 | Postnatal expression of mNaChBac and Kir2.1 using Flpo and F-FLEX switch. **a**, Constitutive overexpression of mNaChBac causes a neuronal migration defect. mNaChBac or mNaChBacMut was overexpressed in a subset of pyramidal cells by *in utero* electroporation of pCAG-mNaChBac-T2A-tdTomato or pCAG-mNaChBacMut-T2A-tdTomato, respectively, on embryonic day 15.5 (E15.5). Representative fluorescent images of coronal sections of V1 obtained at postnatal day 16 or 17 showing that mNaChBac-expressing neurons (left panels) resided not only in layer 2/3, but also in layers 4–6 ($n = 7$ mice), whereas mNaChBacMut-expressing neurons (right panels) are all located in layer 2/3 ($n = 5$ mice). Cortical layers are indicated on the right based on the DAPI staining pattern. **b**, Experimental procedures for conditional expression of mNaChBac or Kir2.1 in a subset of layer 2/3 pyramidal cells. Left, plasmids pAAV-EF1 α -F-FLEX-mNaChBac-T2A-tdTomato or pAAV-EF1 α -F-FLEX-Kir2.1-T2A-tdTomato together with pCAG-EGFP were electroporated *in utero* into V1 on embryonic day 15.5. Successful transfection is indicated by the expression of EGFP. Middle, AAV-hSynapsin-Flpo was injected postnatally into V1. Right, only those neurons that were transfected with either pAAV-EF1 α -F-FLEX-mNaChBac-T2A-tdTomato

or pAAV-EF1 α -F-FLEX-Kir2.1-T2A-tdTomato and infected with AAV-hSynapsin-Flpo expressed mNaChBac-T2A-tdTomato or Kir2.1-T2A-tdTomato, respectively. **c**, Representative fluorescent images of coronal sections of V1 obtained at postnatal day 16 showing that without injection of AAV-hSynapsin-Flpo transfected neurons did not express mNaChBac-T2A-tdTomato (left panels, $n = 2$ mice). The expression of mNaChBac-T2A-tdTomato in transfected neurons was turned on by injection of AAV-hSynapsin-Flpo. These neurons were all properly located in layer 2/3 (right panels, $n = 7$ mice). Cortical layers are indicated on the right based on the DAPI staining pattern. **d**, Schematics of concurrent expression of mNaChBac or Kir2.1 in layer 2/3 pyramidal cells and ChR2 in Pvalb or Sst cells. Plasmids pAAV-EF1 α -F-FLEX-mNaChBac-T2A-tdTomato or pAAV-EF1 α -F-FLEX-Kir2.1-T2A-tdTomato were electroporated *in utero* together with pCAG-EGFP into V1 of *Pvalb-ires-Cre* or *Sst-ires-Cre* mice on embryonic day 15.5. AAV-EF1 α -DIO-hChR2(H134R)-EYFP and AAV-hSynapsin-Flpo were injected postnatally into V1. ChR2 was conditionally expressed in Pvalb or Sst cells, whereas mNaChBac or Kir2.1 was conditionally expressed in a subset of layer 2/3 pyramidal cells.



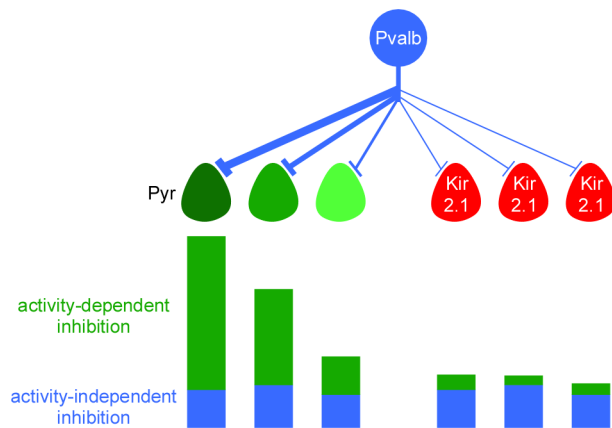
Extended Data Figure 8 | A Flpo recombinase-mediated FLEX (F-FLEX) switch for conditional gene expression. **a**, DNA sequence of the F-FLEX switch cassette. The first *F14* site and *Frt* site were constructed in the forward direction and were separated by a 50-base-pair linker. The second *F14* site and *Frt* site were constructed in the reverse direction and were separated by another 50-base-pair linker. Multiple cloning sites were inserted between the first *Frt* site and the second *F14* site. **b**, Principle of F-FLEX switch. The gene of interest is inserted between the first *Frt* site and the second *F14* site of the F-FLEX switch cassette in an inverted orientation, and is driven by an *EF1 α* promoter. Flp-mediated recombination first occurs between the two *F14* sites or the two *Frt* sites that are in the opposite direction, leading to a reversible inversion of the inverted gene of interest. Flp-mediated recombination then occurs between the two *F14* sites or the two *Frt* sites that are now in the same direction, excising the *Frt* site or the *F14* site between them, respectively. The resulting construct contains only one *F14* site and one *Frt* site, and the gene of

interest is permanently locked in the forward orientation. **c**, Flpo turns on F-FLEX switch. HEK cells were transfected with (1) Flpo, (2) F-FLEX-mNaChBac-T2A-tdTomato, (3) Flpo and F-FLEX-mNaChBac-T2A-tdTomato or (4) Cre and F-FLEX-mNaChBac-T2A-tdTomato. EGFP was co-transfected to monitor the transfection. There was no leaky expression of mNaChBac-T2A-tdTomato in the absence of Flpo. mNaChBac-T2A-tdTomato expression was switched on by the expression of Flpo, but not by Cre. Similar results were obtained with other F-FLEX constructs ($n = 5$). **d**, Flpo does not turn on Cre-dependent DIO switch⁴⁶. HEK cells were transfected with (1) Cre, (2) DIO-hChR2(H134R)-EYFP, (3) Cre and DIO-hChR2(H134R)-EYFP or (4) Flpo and DIO-hChR2(H134R)-EYFP. mRFP was co-transfected to monitor the transfection. There was no leaky expression of hChR2(H134R)-EYFP in the absence of Cre. hChR2(H134R)-EYFP expression was switched on by the expression of Cre, but not by Flpo. Similar results were obtained with other DIO constructs ($n = 2$).



Extended Data Figure 9 | Overexpression of Kir2.1 in a small subset of layer 2/3 pyramidal cells does not affect Pvalb-cell-mediated inhibition onto untransfected pyramidal cells. **a**, Schematic of experiments. Unitary connection from a Pvalb cell onto nearby layer 2/3 pyramidal cells in control mice (left) and onto untransfected pyramidal cells in mice that were electroporated *in utero* with pCAG-Kir2.1-T2A-tdTomato (right). **b**, Connectivity rates from Pvalb cells to layer 2/3 pyramidal cells in control

mice (95%, 57 out of 60) and to untransfected pyramidal cells in electroporated mice (93%, 52 out of 56) are similar ($P = 0.7$). **c**, Cumulative frequencies for uIPSC amplitudes (control: $n = 57$, median, 224.0 pA; untransfected: $n = 52$, median, 190.4 pA; $P = 0.5$). Inset, mean \pm s.e.m. **d**, Summary graph for the average relative deviations of uIPSCs from 20 and 17 similar experiments as in **a**. Bars, mean \pm s.e.m. ($P = 0.6$).



Extended Data Figure 10 | A model for inter-cell variability of Pvalb-cell-mediated inhibition. Schematic illustration of how pyramidal cell activity regulates the inter-cell variability of Pvalb-cell-mediated inhibition. Left, pyramidal cells with different activity levels (dark and light colours indicate high and low activity, respectively) receive different amounts of Pvalb-cell-mediated inhibition (long and short bars indicate more or less inhibition, respectively). Inhibition consists of an activity-dependent component (green bars) and an activity-independent component (blue bars). The activity-dependent component is positively regulated by the pyramidal cell activity and varies accordingly, whereas the activity-independent component is similar across neurons. Right, when the activity of pyramidal cells is suppressed by overexpression of Kir2.1, the activity-dependent component is diminished and the remaining inhibition is largely the activity-independent component. This flooring effect reduces the variability of uIPSC amplitudes among Kir2.1-expressing neurons.

Type I interferon responses in rhesus macaques prevent SIV infection and slow disease progression

Netanya G. Sandler^{1†}, Steven E. Bosinger^{2,3}, Jacob D. Estes⁴, Richard T. R. Zhu¹, Gregory K. Tharp^{2,3}, Eli Boritz¹, Doron Levin⁵, Sathi Wijeyesinghe¹, Krystelle Nganou Makamdop¹, Gregory Q. del Prete⁴, Brenna J. Hill¹, J. Katherina Timmer¹, Emma Reiss¹, Ganit Yarden⁵, Samuel Darko¹, Eduardo Contijoch¹, John Paul Todd⁶, Guido Silvestri², Martha Nason⁷, Robert B. Norgren Jr⁸, Brandon F. Keele⁴, Srinivas Rao⁶, Jerome A. Langer⁹, Jeffrey D. Lifson⁴, Gideon Schreiber⁵ & Daniel C. Douek¹

Inflammation in HIV infection is predictive of non-AIDS morbidity and death¹, higher set point plasma virus load² and virus acquisition³; thus, therapeutic agents are in development to reduce its causes and consequences. However, inflammation may simultaneously confer both detrimental and beneficial effects. This dichotomy is particularly applicable to type I interferons (IFN-I) which, while contributing to innate control of infection^{4–10}, also provide target cells for the virus during acute infection, impair CD4 T-cell recovery, and are associated with disease progression^{6,7,11–19}. Here we manipulated IFN-I signalling in rhesus macaques (*Macaca mulatta*) during simian immunodeficiency virus (SIV) transmission and acute infection with two complementary *in vivo* interventions. We show that blockade of the IFN-I receptor caused reduced antiviral gene expression, increased SIV reservoir size and accelerated CD4 T-cell depletion with progression to AIDS despite decreased T-cell activation. In contrast, IFN- α 2a administration initially upregulated expression of antiviral genes and prevented systemic infection. However, continued IFN- α 2a treatment induced IFN-I desensitization and decreased antiviral gene expression, enabling infection with increased SIV reservoir size and accelerated CD4 T-cell loss. Thus, the timing of IFN-induced innate responses in acute SIV infection profoundly affects overall disease course and outweighs the detrimental consequences of increased immune activation. Yet, the clinical consequences of manipulation of IFN signalling are difficult to predict *in vivo* and therapeutic interventions in human studies should be approached with caution.

We designed and produced an IFN-I receptor antagonist (IFN-1ant) that blocks IFN- α 2 antiviral and antiproliferative activity *in vitro*¹⁹. Six rhesus macaques received 1 mg of IFN-1ant daily for 4 weeks following intrarectal challenge with SIV_{MAC251} (dosage based on previous dose-response studies; Extended Data Fig. 1a–d); nine macaques received saline (Extended Data Fig. 1e). Initial assessment of *in vivo* effects revealed delayed peak mRNA expression of *MX1* and *OAS2* in the IFN-1ant macaques (Extended Data Fig. 2a, b), but peak expression levels did not differ between cohorts. Whole-transcriptome sequencing revealed that expression of most interferon-stimulated genes (ISGs) in peripheral blood mononuclear cells (PBMCs) was significantly decreased at 7 days post-infection (d.p.i.) in the IFN-1ant-treated compared to placebo-treated macaques (Fig. 1a), including the antiviral genes *APOBEC3G* and *MX2*, those that code for cyclic GMP-AMP synthase (cGAS) and tetherin^{4,5,20}, and *IRF7*, a master IFN-I signalling inducer²¹, indicating profound disruption of IFN-I signalling (Fig. 1b). Most ISGs in the IFN-1ant group normalized at 10 and 21 d.p.i. and were upregulated at 28 and 84 d.p.i. (Extended Data Fig. 2c). Consistent with transcriptional data (Extended Data Fig. 2d, e),

APOBEC3G, *TRIM5 α* and *MX2* protein expression by quantitative immunohistochemistry was significantly attenuated in lymph nodes at 4 weeks post-infection (w.p.i.) compared to placebo (Fig. 1c). Thus, IFN-1ant treatment during acute SIV infection resulted in delayed and decreased antiviral gene and protein expression in peripheral blood and lymph nodes.

Consistent with reduced antiviral gene expression, IFN-1ant macaques had significantly higher plasma viral loads (pVLs) than placebo macaques during acute infection (Fig. 2a) and after 20 w.p.i. despite similar numbers of transmitted/founder viruses (measured 10 d.p.i., Extended Data Fig. 8a). Delayed peak ISG expression, however, was predictive of higher pVLs at peak and 12 w.p.i. and higher PBMC-associated SIV *gag* DNA levels at 28 d.p.i. (Extended Data Fig. 2f–h). Additionally, the number of lymph node SIV RNA⁺ cells per mm² as determined by *in situ* hybridization was significantly higher in macaques treated with IFN-1ant compared to placebo during chronic infection (Fig. 2b). Thus, early IFN-I signalling was critical for early and long-term control of SIV replication and virus reservoir size.

Although both groups experienced a similar, significant decrease in circulating CD4 T-cell frequency (Fig. 2c) and CD4/CD8 T-cell ratio (Extended Data Fig. 3a) between 0 and 12 w.p.i., IFN-1ant macaques experienced a profound decline with a lower lymph node CD4 T-cell frequency and CD4/CD8 T-cell ratio beyond 12 w.p.i. (Fig. 2d and Extended Data Fig. 3b). The frequency of CCR5⁺ memory CD4 T cells, potential targets for infection, was significantly lower in blood in IFN-1ant-treated than placebo-treated rhesus macaques through 12 w.p.i. (Fig. 2e), and lymph nodes at 4 and >12 w.p.i. (Fig. 2f), suggesting depletion due to infection. Circulating T-cell activation, reflected by HLA-DR⁺ and Ki67⁺ memory CD4 and CD8 T-cell frequencies, was not significantly different between groups at 4 or >12 w.p.i. (Supplementary Information). However, HLA-DR⁺ and Ki67⁺ memory CD4 and CD8 T-cell frequencies were significantly lower in the lymph nodes of IFN-1ant macaques than placebo at >12 w.p.i. (Extended Data Fig. 3c–f). Taken together, IFN-I signalling blockade during acute SIV infection resulted in attenuated T-cell activation in lymphoid tissue yet accelerated CD4 T-cell depletion.

Clinical outcome ultimately gives the most comprehensive measure of disease state. Consistent with a median life expectancy of 1 year²², the six placebo-treated macaques followed through 44 w.p.i. (three were transferred to another study before 30 w.p.i.) lived, but the IFN-1ant macaques began dying of AIDS at 24 w.p.i. and all were euthanized per protocol for signs of AIDS by 30 w.p.i. (Fig. 2g). Thus, blocking IFN-I signalling during only the first 4 weeks of infection resulted in accelerated disease progression and death from AIDS.

Exploration of the molecular mechanisms underlying the accelerated disease progression by whole-transcriptome sequencing revealed statistically

¹Human Immunology Section, Vaccine Research Center, National Institute of Allergy and Infectious Disease, National Institutes of Health, Bethesda, Maryland 20892, USA. ²Division of Microbiology and Immunology, Emory Vaccine Center, Yerkes National Primate Research Center, Atlanta, Georgia 30322, USA. ³Non-Human Primate Genomics Core, Yerkes National Primate Research Center, Robert W. Woodruff Health Sciences Center, Emory University, Atlanta, Georgia 30322, USA. ⁴AIDS and Cancer Virus Program, Leidos Biomedical Research, Inc., Frederick National Laboratory, Frederick, Maryland 21702, USA. ⁵Department of Biological Chemistry, Weizmann Institute of Science, Rehovot 76100, Israel. ⁶Laboratory of Animal Medicine, Vaccine Research Center, National Institute of Allergy and Infectious Disease, National Institutes of Health, Bethesda, Maryland 20892, USA. ⁷Biostatistics Research Branch, Division of Clinical Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA. ⁸Department of Genetics, Cell Biology and Anatomy, University of Nebraska Medical Center, Omaha, Nebraska 68198, USA. ⁹Department of Pharmacology, Rutgers - Robert Wood Johnson Medical School, Piscataway, New Jersey 08854, USA. [†]Present address: Division of Infectious Diseases, Department of Internal Medicine, University of Texas Medical Branch at Galveston, Galveston, Texas 77555, USA.

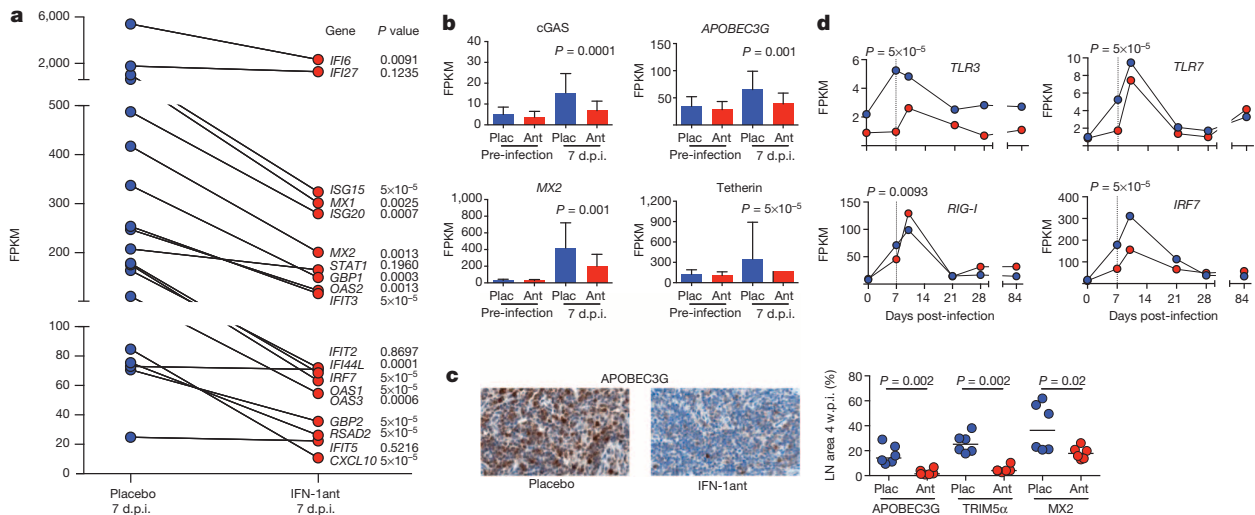


Figure 1 | IFN-1ant suppresses early antiviral responses. **a**, Expression of ISGs in macaques treated with IFN-1ant ($n = 6$) or placebo saline ($n = 9$) 7 days after SIV infection. FPKM (log-transformed fragments per kilobase of transcript per million fragments sequenced) reflects the relative abundance of transcripts. P values indicate differentially expressed genes at 7 d.p.i. **b**, Expression assessed by RNA sequencing (RNA-seq) of antiviral genes APOBEC3G, MX2 and those that code for cGAS and tetherin in PBMCs before and 7 days after SIV infection in macaques that received IFN-1ant (Ant, $n = 6$) or placebo (Plac, $n = 9$) injections. Error bars indicate range. P values were calculated by Mann–Whitney U test. **c**, APOBEC3G, TRIM5 α and MX2

significant enrichment of pathways regulating innate immunity, IFN-I production and T- and B-lymphocyte activation (Extended Data Fig. 4a–c) with significant downregulation of most genes in the IFN-1ant group at 7 d.p.i. compared to placebo-treated controls (Fig. 1d and Extended Data Fig. 2c). Relative to placebo, the most significantly perturbed pathway in the IFN-1ant-treated animals consisted of pathogen-associated pattern recognition receptor (PRR) signalling molecules (Fig. 1d and Extended Data Fig. 4a), with significant downregulation of several viral PRRs (*TLR3*, *TLR7*, *DDX58/RIG-I*, *MDA5/IFIH1*) and their downstream adaptors (*TICAM1/TRIF*) or transcription factors (*IRF7*) in IFN-1ant macaques compared to placebo (Extended Data Fig. 4c). Concordantly, expression of the downstream mediators *IL-6*, *TNF* and *IL-1 β* was significantly reduced.

Consistent with their responsiveness to IFN-I²³, the frequencies of total and cytotoxic CD16⁺ natural killer (NK) cells were significantly lower in the IFN-1ant group than placebo at >12 w.p.i., although there were no differences at 4 w.p.i. (Extended Data Fig. 3g–i). However, we observed no significant differences in phenotype, function or timing of CD4 or CD8 T-cell responses (Extended Data Fig. 5 and Supplementary Information).

Collectively, these data suggest that IFN-I signalling early in SIV infection is critical for innate immune control of virus replication and that its antagonism, even if only brief during the acute phase, results in decreased virus control, accelerated CD4 T-cell depletion and progression to AIDS.

Given these findings, we hypothesized that administering IFN-I could improve SIV_{MAC251} control despite evidence suggesting that inflammation exacerbates virus acquisition and disease progression^{1–3,13}. Six macaques received 6 $\mu\text{g kg}^{-1}$ pegylated IFN- α 2a dosed weekly, as determined in prior studies²⁴, starting 1 week before challenge and continued through 4 weeks after systemic infection (defined as detectable pVL 7 days post-challenge). Macaques were followed until 12 w.p.i. then euthanized per protocol (Extended Data Fig. 1e). Whereas all nine placebo macaques were infected after the first intrarectal inoculation, IFN- α 2a treatment significantly delayed systemic infection, necessitating two, three or five challenges to infect these macaques (Fig. 3a), and significantly decreased the number of transmitted/founder variants (Extended Data Fig. 8a). The macaques that required more challenges had fewer transmitted/founder variants (Fig. 3b); however, the circulating viruses at peak viral load in both groups

were equally susceptible to *in vitro* IFN α inhibition (data not shown). Thus, treatment with IFN- α 2a during SIV challenge increased host resistance to systemic infection.

We assessed whether ISGs may have contributed to resistance to infection. We found that *MX1* and *OAS2* gene expression detected by quantitative reverse transcription PCR (qRT–PCR) increased after one IFN- α 2a dose but decreased after repeated administration, suggesting an IFN-desensitized state (Extended Data Fig. 6a–d) resulting in no significant differences between treated and placebo groups on the day of infectious challenge. Furthermore, IFN- α 2a macaques had lower ISG expression compared to placebo at 7 and 10 d.p.i., including those encoding cGAS, APOBEC3G, MX2 and tetherin (Fig. 3c–f and Extended Data Fig. 7a–b). At 4 w.p.i., after 6, 7 or 9 doses of IFN- α 2a, lymph node TRIM5 α protein expression was significantly lower in the IFN- α 2a group compared to placebo (Extended Data Fig. 8b). To explore the effects of IFN- α 2a administration on ISG expression further, we performed whole-transcriptome sequencing on uninfected rhesus macaques administered pegylated IFN- α 2a for 3 weeks. Seven days after one IFN- α 2a dose, expression of virus restriction factors including *TRIM22*, *MX2* and *IRF7* was increased in PBMCs, lymph nodes and rectum (Extended Data Fig. 7c–h). However, after 3 doses these ISGs returned to pre-treatment expression levels (*MX2*) or lower (*TRIM22* and *IRF7*), consistent with the timing of infection of the SIV-challenged macaques. (Extended Data Fig. 7c–h). We found no *in vitro* IFN-neutralizing activity in the plasma at the time of infection to explain the loss of exogenous IFN- α 2a activity (Extended Data Fig. 6e and Supplementary Information), suggesting that ISG downregulation probably occurred as a result of intrinsic regulatory mechanisms. Indeed, *FOXO3a*, a central regulator of IFN-I feedback²⁵, was significantly upregulated in the IFN- α 2a compared to placebo macaques at 7 d.p.i. (Fig. 3g and Extended Data Fig. 6g) and, concordantly, ISGs demonstrated to be repressibly bound by FOXO3a (ref. 25) had lower expression at 7 d.p.i. (Extended Data Fig. 6h). We also assessed a broad panel of genes predicted to be bound by FOXO3a (ref. 25) using gene-set enrichment analysis and observed a significantly lower cumulative ranking of FOXO3a targets in the 7 d.p.i. IFN- α 2a-treated macaque transcriptome data when contrasted to the 7 d.p.i. placebo data (Fig. 3h), suggesting increased FOXO3a-mediated repression in the IFN- α 2a macaques. Indeed, after 21 days of

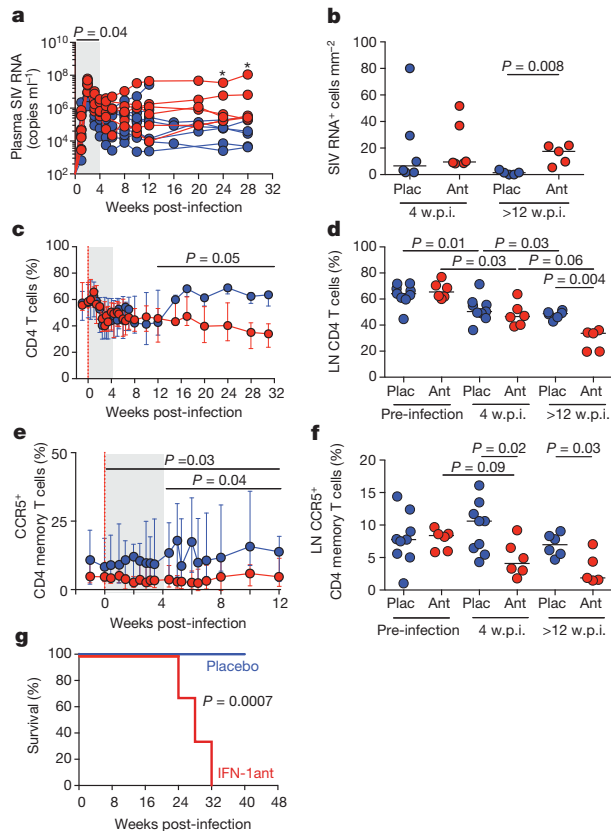


Figure 2 | IFN-1ant accelerates disease progression in SIV-infected rhesus macaques. **a**, Plasma SIV RNA levels during acute and chronic SIV infection in macaques treated with IFN-1ant ($n = 6$) or placebo saline ($n = 9$). * $P < 0.05$. Shading indicates treatment period. P value represents the comparison between groups of the areas under the curve (AUC) (0–4 w.p.i.). **b**, SIV RNA-containing cells in the lymph nodes by *in situ* hybridization at 4 and 12 w.p.i. in IFN-1ant (Ant, $n = 6$) and placebo (Plac, $n = 6$) macaques. Horizontal bars represent median values. P value was calculated by Mann–Whitney U test. **c**, Frequency of CD4 T cells in peripheral blood during acute and chronic SIV infection in macaques treated with IFN-1ant ($n = 6$) or placebo saline ($n = 9$). Error bars indicate range. Red vertical line indicates day 0 of systemic SIV infection. Shading indicates treatment period. P value represents the comparison between groups of the AUC (12–32 w.p.i.). **d**, Frequency of CD4 T cells in lymph nodes before SIV infection and at 4 or >12 w.p.i. for IFN-1ant (Ant, $n = 6$) and placebo (Plac, $n = 9$) macaques. Horizontal bars represent median values. P values at different time points within treatment groups were calculated by Wilcoxon matched pairs signed rank test and between groups by Mann–Whitney U test. **e**, Frequency of CCR5⁺ memory (CD28⁺CD95⁺ or CD28[−]CD95^{+/−}) CD4 T cells in peripheral blood in macaques treated with IFN-1ant ($n = 6$) or placebo saline ($n = 9$). Error bars indicate range. Red vertical line indicates day 0 of systemic SIV infection. Shading indicates treatment period. P values represent the comparison between groups of the AUC (0–12 w.p.i. and 4–12 w.p.i.). **f**, Frequency of CCR5⁺ memory CD4 T cells in lymph nodes in macaques treated with IFN-1ant ($n = 6$) or placebo saline ($n = 9$). Horizontal bars represent median values. P values at different time points within treatment groups were calculated by Wilcoxon matched pairs signed rank test and between groups by Mann–Whitney U test. **g**, Kaplan–Meier survival curve comparing macaques treated with IFN-1ant ($n = 6$) to macaques that received placebo ($n = 9$). P value indicates the significance by logrank (Mantel–Cox) test for survival by 32 w.p.i. For all panels, IFN-1ant-treated macaques are represented in red, placebo-treated macaques in blue.

IFN- α 2a treatment in the unchallenged macaques, increased *FOXO3a* expression was associated with ISG downregulation (Extended Data Fig. 6f). While these data do not exclude additional mechanisms, they suggest that ISG downregulation in the IFN- α 2a-treated macaques was a consequence of endogenous homeostatic control rather than neutralization of exogenous IFN- α 2a. Thus, exogenous augmentation of IFN-I signalling was associated with enhanced protection against SIV acquisition, but susceptibility to exacerbated systemic infection once ISG expression waned.

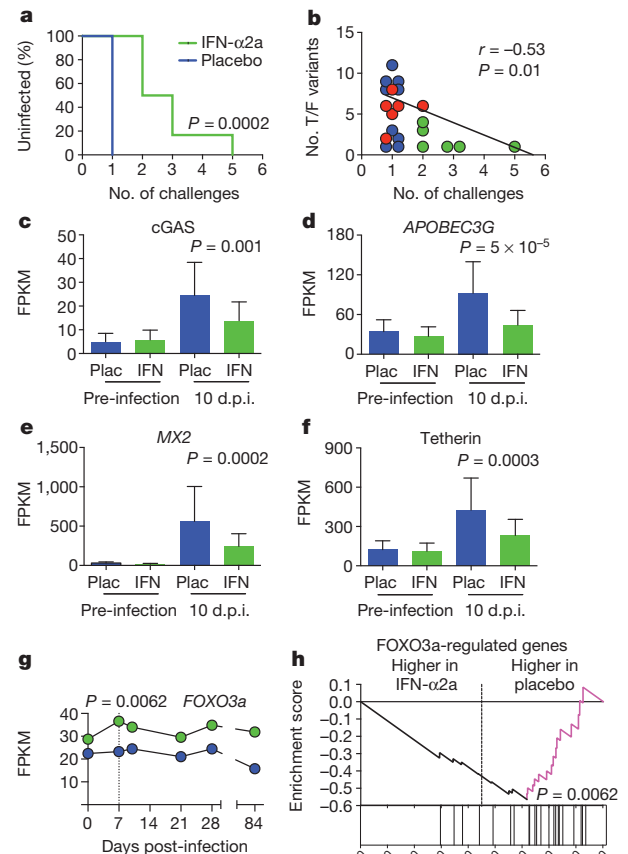


Figure 3 | IFN- α 2a treatment transiently prevents systemic infection but results in an IFN-tolerant state. **a**, Kaplan–Meier survival curve comparing the number of SIV_{MAC251} rectal challenges required to achieve systemic infection in macaques treated with IFN- α 2a ($n = 6$) or placebo saline ($n = 9$). P value indicates the significance by logrank (Mantel–Cox) test of the number of challenges required for systemic infection, between 1 and 5 challenges. **b**, Correlation between the number of challenges needed to achieve systemic infection and the number of transmitted/founder (T/F) variants in IFN- α 2a (green, $n = 6$), IFN-1ant (red, $n = 6$) and placebo (blue, $n = 9$) macaques. P value indicates the significance of the correlation between the number of challenges and the number of T/F variants in all groups. r indicates the Spearman’s rank correlation coefficient. **c–f**, Expression of antiviral mediators in PBMCs in IFN- α 2a-treated (IFN, $n = 6$) macaques compared to placebo (Plac, $n = 9$) at 10 d.p.i. Error bars indicate range. P values represent the comparison of FPKMs between IFN- α 2a and placebo at 10 d.p.i. by Mann–Whitney U test. **g**, Expression profile of *FOXO3a*, a negative regulator of type I IFN signalling. P value represents the comparison of *FOXO3a* FPKM between IFN- α 2a ($n = 6$) and placebo ($n = 9$) macaques at 7 d.p.i. **h**, Gene-set enrichment analysis in IFN- α 2a ($n = 6$) and placebo ($n = 9$) macaques of genes previously demonstrated to be overexpressed in *FOXO3*^{−/−} macrophages²⁵. The line plot indicates the running-sum of the enrichment score; the leading edge is indicated in magenta. The relative positions of all genes within the ranked data set are shown in the stick plot below the x axis. P value indicates statistical significance of the enrichment score, reflecting lower cumulative ranking of *FOXO3a* targets in IFN- α 2a-treated macaques compared to placebo at 7 d.p.i. For all panels, IFN- α 2a-treated macaques are represented in green, placebo-treated macaques in blue.

Given their roles in virus control, T and NK cells were evaluated. No changes in SIV-specific CD4 or CD8 T-cell responses developed with repeated challenges, suggesting that resistance to infection did not depend on adequate T-cell responses (Supplementary Information). However, higher circulating CD56⁺ NK-cell frequencies after starting IFN- α 2a predicted more challenges necessary for infection and were associated with resistance to infection (Extended Data Fig. 8c). Once the CD56⁺ NK-cell

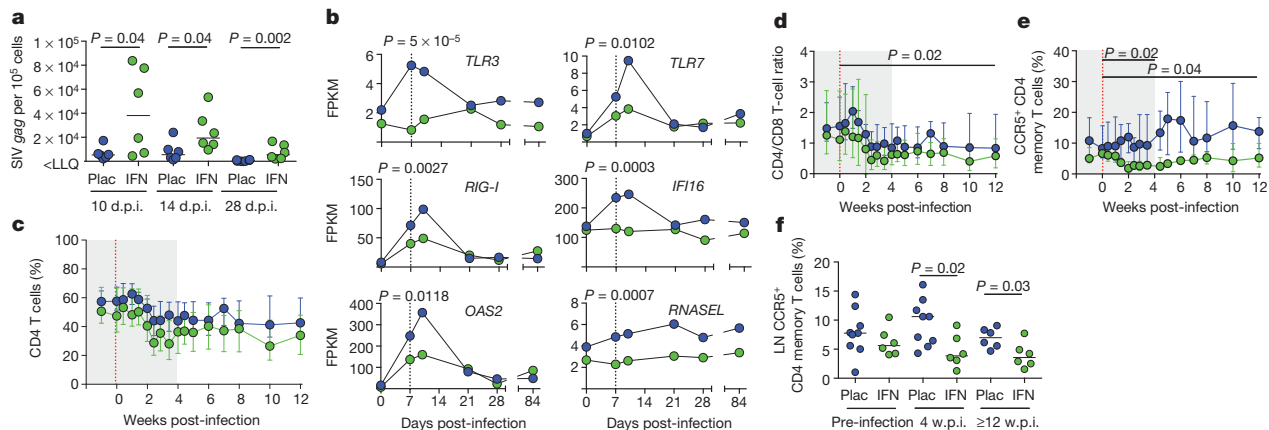


Figure 4 | IFN- α 2a accelerates disease progression. **a**, PBMC-associated SIV *gag* DNA at 10, 14 and 28 d.p.i. in IFN- α 2a macaques (IFN, $n = 6$) and placebo (Plac, $n = 6$) macaques. LLQ indicates lower limit of quantification. Horizontal bars represent median values. P values were calculated by Mann–Whitney U test. **b**, Expression of differentially expressed genes involved in pattern recognition receptor signalling. P values represent the comparison between FPKMs of IFN- α 2a ($n = 6$) and placebo ($n = 9$) macaques at 7 d.p.i. **c**, Frequency of CD4 T cells in peripheral blood during acute and early SIV infection in IFN- α 2a ($n = 6$) and placebo ($n = 9$) macaques. Error bars indicate range. Shading indicates treatment period. Red vertical line indicates day 0 of systemic SIV infection. **d**, CD4/CD8 T-cell ratio in peripheral blood during acute and early SIV infection in IFN- α 2a ($n = 6$) and placebo ($n = 9$)

macaques. Error bars indicate range. Shading indicates treatment period. Red vertical line indicates day 0 of systemic SIV infection. P value represents the comparison between groups of the AUC (0–12 w.p.i.). **e**, Frequency of CCR5⁺ memory CD4 T cells in peripheral blood in IFN- α 2a ($n = 6$) and placebo ($n = 9$) macaques. Error bars indicate range. Shading indicates treatment period. Red vertical line indicates day 0 of systemic SIV infection. P values represent the comparison between groups of the AUC (0–4 and 0–12 w.p.i.). **f**, Frequency of CCR5⁺ memory CD4 T cells in lymph nodes in IFN- α 2a ($n = 6$) and placebo ($n = 9$) macaques. Horizontal bars represent median values. P values were calculated by Mann–Whitney U test. For all panels, IFN- α 2a-treated macaques are represented in green, placebo-treated macaques in blue.

frequency declined, the macaques became infected. In rectal biopsies at 28 d.p.i., more resistant macaques had higher CD16⁺ NK-cell frequencies (Extended Data Fig. 8d). Together, these data suggest that IFN- α 2a-induced innate immunity, rather than T-cell responses, protected against SIV infection.

Despite fewer transmitted/founder variants, pVLs did not differ significantly between IFN- α 2a and placebo groups (Extended Data Fig. 8e), potentially obscured by the variability and small number of macaques. PBMC-associated SIV *gag* DNA levels, however, were significantly higher in the IFN- α 2a group than placebo at 10, 14 and 28 d.p.i. (Fig. 4a). While the circulating CD4 T-cell frequency (Fig. 4c) and CD4/CD8 T-cell ratio (Fig. 4d) declined between 0 and 4 w.p.i. in both groups, the CD4/CD8 T-cell ratio was significantly lower in the IFN- α 2a group (based on area under the curve (AUC) (0–12 w.p.i.)). The CCR5⁺ memory CD4 T-cell frequency in blood was significantly lower in IFN- α 2a than placebo macaques during acute (AUC(0–4 w.p.i.)) (Fig. 4e) and chronic infection (AUC(0–12 w.p.i.)) and in lymph nodes at 4 (Fig. 4f) and 12 w.p.i., although the frequency of CD4 T cells and CD4/CD8 T-cell ratio in lymph nodes and jejunum were similar between groups at 4 and ≥ 12 w.p.i. (data not shown). Thus, SIV-infected IFN- α 2a-treated macaques had increased CD4 T-cell-associated virus load and greater CD4 T-cell loss with preferential depletion of the CCR5⁺ subset.

We assessed whether increased immune activation was associated with this CD4 T-cell loss. The circulating Ki67⁺ memory CD4 and CD8 T-cell frequencies were lower in the IFN- α 2a compared to placebo group during acute infection only (AUC(0–4 w.p.i.)) with no differences in the frequencies of HLA-DR⁺ memory T cells (Extended Data Fig. 9a–d). In the lymph nodes, the frequencies of Ki67⁺ memory CD4 T cells at 4 w.p.i. and HLA-DR⁺ memory CD4 T cells at ≥ 12 w.p.i. were significantly lower in the IFN- α 2a than placebo group with no differences in CD8 T cells (Extended Data Fig. 9e–h). Thus, the IFN- α 2a macaques had similar or less immune activation compared to placebo macaques.

We further explored the mechanisms underlying the increased cell-associated SIV and CD4 T-cell depletion in IFN- α 2a macaques using whole-transcriptome sequencing of PBMCs. As with IFN-Iant, IFN- α 2a administration significantly affected the PRR signalling pathway (Extended Data Fig. 10a). *C1q*, *TLR3*, *TLR7* and *RIG-I* were downregulated in the IFN- α 2a macaques, yet expression of *IL-6*, *TNF* and *IL-1 β* was increased at 7 d.p.i. (Fig. 4b and Extended Data Fig. 10b). Mediators

of stress responses and cell survival downstream of IL-6 signalling, such as *MAP2K3*, and other anti-apoptotic genes such as *DIABLO* and *BCL2L1* were upregulated at 7 d.p.i. (Extended Data Fig. 10b, 10c), whereas pro-apoptotic genes including *CASP10* were downregulated. Thus, early IFN- α 2a administration increased early expression of proinflammatory cytokines despite decreased PRR expression and delayed induction of apoptotic pathways.

We next assessed T- and NK-cell-mediated immunity. IFN- α 2a-treated macaques had intact and even enhanced SIV-specific CD8 T-cell responses with no deficits in CD4 T-cell responses (Extended Data Fig. 8f–i and Supplementary Information). Whereas there were no differences at 4 w.p.i., between 4 and 12 w.p.i., the CD56⁺ NK-cell frequency increased and CD16⁺ NK-cell frequency decreased in the IFN- α 2a but not the placebo group. The frequencies of CD16⁺ (Extended Data Fig. 8j), CD107a⁺ and granzyme B⁺ NK cells at 12 w.p.i. were subsequently lower in the IFN- α 2a-treated macaques compared to placebo.

Taken together, these data show that despite initially increasing ISG expression and conferring resistance to SIV infection, continued IFN- α 2a treatment resulted in an IFN-desensitized state, with decreased antiviral gene expression, increased susceptibility to infection, increased cell-associated virus load and greater CD4 T-cell depletion compared to placebo.

Thus, both IFN-I receptor blockade and IFN- α 2a administration ultimately resulted in decreased and delayed IFN-I responses, the biological outcomes of which reveal a pivotal role for IFN-I signalling in acute retroviral exposure in primates that overshadows the potential harm of increased inflammation. That a delay of as few as 3 days in antiviral gene expression resulted in accelerated disease progression suggests that SIV disease course is determined very early and depends upon the precise timing of peak antiviral activity. The observation of shortened time-to-death despite the eventual normalization of ISG expression during chronic infection clearly shows that a resurgence of late antiviral activity cannot compensate for compromised early virus control. Indeed, administration of IFN- α in chronic HIV infection has given inconsistent results for virus load and CD4 T-cell counts with no effect on disease outcome^{7–9}. An interesting parallel to our study was recently described in mice with chronic lymphocytic choriomeningitis virus infection where persistent IFN-I signalling exerts antiviral effects but also leads to immune hyperactivation and suppression of antiviral

T-cell responses^{14,15}. Our findings differ from a previous study that reported unchanged virus burden with IFN- α 2a administration during chronic SIV infection²⁴ and a study of IFN- α 2b or the IFN- α B/D chimera in rhesus macaques challenged intravenously with highly pathogenic SIV DeltaB670 that showed no protection from infection but decreased peak antigenaemia²⁶. The initial protection from infection and decrease in transmitted/founder variants observed in our study highlight the rectal mucosa's function as a barrier to SIV during transmission in the context of ISG induction. It is tempting to speculate that administration of IFN- α 2a before SIV challenge also facilitated NK-cell activation and recruitment to the rectum and contributed to protection from infection²⁷. Given the relative IFN-I resistance of transmitted/founder viruses^{28,29} and the induction of an IFN-desensitized state associated with the upregulation of the IFN-I pathway repressor *FOXO3a* (ref. 25), our findings add a cautionary note to adjuvanted HIV vaccines or other prevention approaches that induce ISGs at mucosal surfaces. Furthermore, while the rectum contains many resident target CD4 T cells³⁰ that antiviral mediators can protect, sites with few resident target cells, such as the female genital tract¹⁶, may depend on IFN-I signalling for CD4 T-cell recruitment and virus propagation. Thus, inflammation might attenuate transmission at the former site but exacerbate it at the latter³. In conclusion, disease progression in HIV infection emerges from the balance between the beneficial antiviral effects of inflammation, its detrimental systemic and immunologic effects, and its unique role of providing activated CD4 T-cell targets for HIV. Interfering with one part of this unstable equilibrium has unpredictable consequences. Thus, while there is good reason to use both pro- and anti-inflammatory therapeutic approaches in the treatment and prevention of HIV infection, they should be embarked upon with careful assessment of the virological and immunological consequences before widespread implementation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 11 April; accepted 4 June 2014.

Published online 9 July 2014.

- Hunt, P. W. *et al.* Gut epithelial barrier dysfunction and innate immune activation predict mortality in treated HIV infection. *J. Infect. Dis.* <http://dx.doi.org/10.1093/infdis/jiu238> (21 April 2014).
- Roberts, L. *et al.* Genital tract inflammation during early HIV-1 infection predicts higher plasma viral load set point in women. *J. Infect. Dis.* **205**, 194–203 (2012).
- Naranbhai, V. *et al.* Innate immune activation enhances HIV acquisition in women, diminishing the effectiveness of tenofovir microbicide gel. *J. Infect. Dis.* **206**, 993–1001 (2012).
- Schoggins, J. W. *et al.* A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* **472**, 481–485 (2011).
- Schoggins, J. W. *et al.* Pan-viral specificity of IFN-induced genes reveals new roles for cGAS in innate immunity. *Nature* **505**, 691–695 (2013).
- Gonzalez-Navajas, J. M., Lee, J., David, M. & Raz, E. Immunomodulatory functions of type I interferons. *Nature Rev. Immunol.* **12**, 125–135 (2012).
- Lane, H. C. *et al.* Anti-retroviral effects of interferon- α in AIDS-associated Kaposi's sarcoma. *Lancet* **332**, 1218–1222 (1988).
- Manion, M. *et al.* Interferon-alpha administration enhances CD8+ T cell activation in HIV infection. *PLoS ONE* **7**, e30306 (2012).
- Azzoni, L. *et al.* Pegylated Interferon alfa-2a monotherapy results in suppression of HIV type 1 replication and decreased cell-associated HIV DNA integration. *J. Infect. Dis.* **207**, 213–222 (2013).
- Feld, J. J. & Hoofnagle, J. H. Mechanism of action of interferon and ribavirin in treatment of hepatitis C. *Nature* **436**, 967–972 (2005).
- Stacey, A. R. *et al.* Induction of a striking systemic cytokine cascade prior to peak viremia in acute human immunodeficiency virus type 1 infection, in contrast to more modest and delayed responses in acute hepatitis B and C virus infections. *J. Virol.* **83**, 3719–3733 (2009).
- Fraietta, J. A. *et al.* Type I interferon upregulates Bak and contributes to T cell loss during human immunodeficiency virus (HIV) infection. *PLoS Pathog.* **9**, e1003658 (2013).
- Abel, K. *et al.* The relationship between simian immunodeficiency virus RNA levels and the mRNA levels of α/β interferons (IFN- α /beta) and IFN- α / β -inducible Mx in lymphoid tissues of rhesus macaques during acute and chronic infection. *J. Virol.* **76**, 8433–8445 (2002).
- Tejaro, J. R. *et al.* Persistent LCMV infection is controlled by blockade of type I interferon signaling. *Science* **340**, 207–211 (2013).
- Wilson, E. B. *et al.* Blockade of chronic type I interferon signaling to control persistent LCMV infection. *Science* **340**, 202–207 (2013).
- Li, Q. *et al.* Glycerol monolaurate prevents mucosal SIV transmission. *Nature* **458**, 1034–1038 (2009).
- Haas, D. W. *et al.* A randomized trial of interferon alpha therapy for HIV type 1 infection. *AIDS Res. Hum. Retroviruses* **16**, 183–190 (2000).
- Fernandez, S. *et al.* CD4+ T-cell deficiency in HIV patients responding to antiretroviral therapy is associated with increased expression of interferon-stimulated genes in CD4+ T cells. *J. Infect. Dis.* **204**, 1927–1935 (2011).
- Levin, D. *et al.* Multifaceted activities of type I interferon are revealed by a receptor antagonist. *Sci. Signal.* **7**, ra50 (2014).
- Goujon, C. *et al.* Human MX2 is an interferon-induced post-entry inhibitor of HIV-1 infection. *Nature* **502**, 559–562 (2013).
- Honda, K. *et al.* IRF-7 is the master regulator of type-I interferon-dependent immune responses. *Nature* **434**, 772–777 (2005).
- Barouch, D. H. *et al.* Vaccine protection against acquisition of neutralization-resistant SIV challenges in rhesus monkeys. *Nature* **482**, 89–93 (2012).
- Carrington, M. & Alter, G. Innate immune control of HIV. *Cold Spring Harb. Perspect. Med.* **2**, a007070 (2012).
- Asmuth, D. M. *et al.* Pegylated interferon- α 2a treatment of chronic SIV-infected macaques. *J. Med. Primatol.* **37**, 26–30 (2008).
- Litvak, V. *et al.* A FOXO3-IRF7 gene regulatory circuit limits inflammatory sequelae of antiviral responses. *Nature* **490**, 421–425 (2012).
- Schellekens, H. *et al.* The effect of recombinant human interferon α B/D compared to interferon α 2b on SIV infection in rhesus macaques. *Antiviral Res.* **32**, 1–8 (1996).
- Waggoner, S. N., Daniels, K. A. & Welsh, R. M. Therapeutic depletion of natural killer cells controls persistent infection. *J. Virol.* **88**, 1953–1960 (2014).
- Parrish, N. F. *et al.* Phenotypic properties of transmitted founder HIV-1. *Proc. Natl Acad. Sci. USA* **110**, 6626–6633 (2013).
- Fenton-May, A. E. *et al.* Relative resistance of HIV-1 founder viruses to control by interferon-alpha. *Retrovirology* **10**, 146 (2013).
- McElrath, M. J. *et al.* Comprehensive assessment of HIV target cells in the distal human gut suggests increasing HIV susceptibility toward the anus. *J. Acquir. Immune Defic. Syndr.* **63**, 263–271 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to acknowledge A. Zimin for his work in creating the MuSuRCA rhesus assembly, C. Miller for the gift of 6 rhesus macaques and Y. Peleg and S. Albeck at the Israel Structure Proteomic Center and G. Jona from Weizmann Institute Biological services for helping with protein production and purification; A. Roque and N. Haining for initial work on the pilot study; N. Modi, D. Ambrozak, R. Koup, M. Ghosh, I. Srivastava, R. Schwartz, F. Villinger, K. Zoon, J. Bekisz, K. Ghneim, A. Filali, R. Sekaly, L. Mach and L. Shen for their assistance on the current project; and A. Somasunderam for additional support. This project was supported by NIH Intramural Funding, federal funds from NCI/NIH Contract HHSN261200800001E, NIH R24 RR017444, NIH AI-076174, I-CORE Program of the Planning and Budgeting Committee and the Israel Science Foundation grant No. 1775/12.

Author Contributions N.G.S. designed and coordinated the study, developed and performed experiments, interpreted the data and prepared the manuscript. S.E.B. analysed and interpreted the sequencing data, generated figures and contributed to manuscript preparation. J.D.E. contributed to study design and developed and performed *in situ* hybridization and immunohistochemistry assays. R.T.R.Z. processed samples, performed flow cytometry and analysis, performed qRT-PCR and generated the sequencing libraries. G.K.T. analysed and interpreted the sequencing data and generated figures. E.B. developed the library generation protocol and supervised library generation. D.L. and G.Y. synthesized the IFN-1ant. S.W. generated sequencing libraries and assisted in analysis of the sequencing data. K.N.M. assisted with sample processing, performed flow cytometry assays, and assessed plasma for neutralizing activity. G.Q.D.P. evaluated circulating SIV for IFN resistance. B.J.H. designed, performed and analysed qRT-PCR assays. J.K.T. processed samples and performed ELISAs. E.R. assisted with sample processing and performed flow cytometry assays. S.D. assisted with sequencing analysis. E.C. assisted with sample processing and performed flow cytometry assays. J.P.T. performed SIV inoculations and coordinated the study at Bioqual. G.S. established the Non-Human Primate Sequencing Core and facilitated sequencing analysis and contributed to data interpretation. M.N. assisted with statistical analyses. R.B.N. generated the MuSuRCA *Macaca mulatta* assembly. B.F.K. sequenced the transmitted/founder variants. S.R. contributed to study design and followed the rhesus macaques clinically. J.A.L. contributed to IFN-1ant design and assisted with analysis. J.D.L. contributed to study design, assessment for IFN-resistant viruses and manuscript preparation. G.Sc. contributed to study design, IFN-1ant design and production and assisted with analysis. D.C.D. designed and supervised the study, interpreted the data and prepared the manuscript.

Author Information Gene expression data are available at the Gene Expression Omnibus under accession codes GSM1298835 through GSM1299037. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.C.D. (ddouek@mail.nih.gov).

METHODS

Dose escalation study. We performed a dose escalation study in two rhesus macaques with chronic SIV_{MAC251} infection and different frequencies of CD4 T cells and CCR5⁺ CD4 and CD8 T cells to maximize the likelihood of detecting a response. IFN-1ant was dosed three times a week based on dosing of recombinant IFN- α 2a (Roferon-A, Roche, Switzerland) for hepatitis C infection. We administered 50 μ g of IFN-1ant for one week, 200 μ g for one week, 500 μ g for one week and 800 μ g for one week. Based on a dose-dependent increase in CD4 T-cell frequency and decrease in the frequencies of CCR5⁺ CD4 and CCR5⁺ and Ki67⁺ CD8 T cells (Extended Data Fig. 1a–d), and because of ease of administration, we decided on 1 mg of IFN-1ant. Daily dosing was chosen as the effects appeared to be variable based on the time since last dosing.

Macaques and experimental design. To examine the effects of blocking type I IFN signalling, healthy, SIV-uninfected Mamu A01[−]B08[−]B17[−] adult *Macaca mulatta* received 1 mg daily intramuscularly of the IFN-I receptor antagonist (IFN-1ant, 1 mg ml^{−1}, $n = 6$; synthesized as previously described¹⁹) or 1 ml normal saline ($n = 9$) by intramuscular injection for 4 weeks (see Extended Data Fig. 1e). The macaques were inoculated weekly, starting the first day of treatment, with 1 ml of SIV_{MAC251} via rectal challenge (1 ml of a 1:25 dilution, stock 3×10^8 SIV RNA copies ml^{−1}), up to three times until infection was confirmed (SIV pVL > 250 copies ml^{−1}). Blood was sampled several times per week, and lymph node and jejunal biopsies were performed before treatment and at the end of the 4-week period. All blood draws and biopsies were performed before inoculation with SIV_{MAC251} and before any drug administration on that day. The macaques were followed for up to 40 w.p.i.

To examine the effect of exogenous IFN-I treatment, six healthy, SIV-uninfected Mamu A01[−]B08[−]B17[−] adult macaques received 6 μ g kg^{−1} pegylated IFN- α 2a (Pegasys, Genentech USA) intramuscularly weekly starting one week before the first SIV_{MAC251} inoculation. Dose was based on prior efficacy studies in rhesus macaques²⁴. The macaques were challenged intrarectally weekly on the same day as, but before, IFN- α 2a administration, with the same high-dose SIV_{MAC251} inoculation as the IFN-1ant and placebo macaques, until infection was confirmed. All blood draws and biopsies were performed before inoculation with SIV_{MAC251} and before any drug administration on that day. The macaques were followed for a total of 12 weeks after infection and then euthanized per protocol. To examine the effect of exogenous IFN-I treatment in the absence of SIV infection, three healthy, SIV-uninfected Mamu A01[−]B08[−]B17[−] adult macaques received 6 mg kg^{−1} pegylated IFN- α 2a (Pegasys, Genentech USA) intramuscularly weekly for 3 weeks. Blood was sampled twice weekly starting 1 week before IFN- α 2a treatment through to 1 week after the last dose. The lymph node, jejunum and rectum were biopsied before IFN- α 2a treatment, 1 week after the first dose and 1 week after the last dose.

All macaques were housed at Bioqual, Inc., and assigned randomly to treatment or placebo arms. TRIMCyp, the fusion protein derived from TRIM5 and cyclophilin A, was present in 1 placebo macaque, 2 IFN-1ant macaques and 0 IFN- α 2a macaques. One TRIM5 α SPRY deletion was present in 4 placebo macaques, 3 IFN-1ant macaques and 1 IFN- α 2a macaque (the macaque that required 5 challenges to become systemically infected) and two TRIM5 α SPRY deletions in 2 IFN-1ant macaques. There was no significant difference in genotype distribution between placebo and IFN-1ant macaques or between placebo and IFN- α 2a macaques based on Fisher's test. Based on availability, the 6 IFN-1ant macaques and 9 placebo macaques were male and ages 4 to 7 years, and the 6 IFN- α 2a macaques were female with ages 9 to 15 years. The Vaccine Research Center Animal Care and Use Committee approved all study protocols and procedures.

Samples. Blood was collected in EDTA tubes. Plasma was collected and PBMCs were isolated by Ficoll density centrifugation. All of the jejunum and half the lymph node biopsy tissues and half of every tissue collected at necropsy were placed in RPMI with 10% fetal bovine serum (FBS) and transported on wet ice. The remaining tissues were placed in 4% paraformaldehyde and kept at room temperature overnight before being transferred to 80% ethanol and stored at 4 °C. Tissues were subsequently paraffin-embedded for *in situ* hybridization and immunohistochemistry. Intestinal samples were incubated with RPMI + collagenase D (1 mg ml^{−1}) (Roche) + Penicillin-Streptomycin-Glutamine (Gibco) at 37 °C for 30 min and then passed through a 70 μ m filter. Lymph node samples were passed through a 70 μ m filter to remove debris. Cells isolated from both peripheral blood and tissues were either stained for flow cytometry or cryopreserved.

Flow cytometry. Cellular activation and cell cycle entry were assessed by flow cytometry. PBMCs and cells from the jejunum and lymph nodes were stained with Aqua LIVE/DEAD Fixable Dead Cell Stain and antibodies to the following: CD4 Qd605, CD8 Qd655 (Invitrogen); CD3 APC-Cy7, CD95 PE-Cy5, CD14 Pacific Blue, CCR5 PE, CCR7 PE-Cy7, Ki67 FITC (BD Biosciences); CD28 ECD (Beckman Coulter); and HLA-DR Alexa 700PE (BD Biosciences, in-house conjugate). Cells were permeabilized using a Cytotfix/Cytoperm kit (BD Biosciences) for Ki67 detection and fixed with 1% formaldehyde (Tousimis).

To assess antigen-specific responses and NK cell subsets, cryopreserved PBMCs were stimulated for 6 h at 37 °C with SIV Gag/Env/Pol peptide pool (2 μ g ml^{−1}) in the presence of CD28 and CD49d (BD Biosciences) and brefeldin A (Sigma, 10 μ g ml^{−1}). Cells were stained with Aqua LIVE/DEAD Fixable Dead Cell Stain (Invitrogen) and antibodies to the following: CD3 APC-Cy7, CD14 Pacific Blue, IFN- γ Cy7PE, TNF APC, CD107a FITC (BD Biosciences); Granzyme B PE (Caltag); CD4 Qd605, CD8 Qd655 (Invitrogen); CD16 Ax594, CD20 Ax700PE, CD56 Cy5PE, Perforin Ax680 (in-house conjugates, BD Biosciences). Cells were permeabilized using a Cytotfix/Cytoperm kit (BD Biosciences) for intracellular cytokine detection and fixed with 1% formaldehyde (Tousimis).

To evaluate cellular exhaustion, cryopreserved PBMCs were stimulated for 6 h at 37 °C with SIV Gag/Env/Pol peptide pool (2 μ g ml^{−1}) in the presence of brefeldin A (10 μ g ml^{−1}). Cells were stained with Aqua LIVE/DEAD Fixable Dead Cell Stain (Invitrogen) and antibodies to the following: CD3 APC-Cy7, CD95 Cy5PE, IFN- γ FITC, TNF APC, Bcl-2 PE (BD Biosciences); CD28 ECD (Beckman Coulter); ICOS Pacific Blue (Biolegend); CD4 Qd605, CD8 Qd655, Streptavidin Cy7PE (Invitrogen); PD-1 biotinylated (R&D Systems). Cells were permeabilized using a Cytotfix/Cytoperm kit (BD Biosciences) for intracellular cytokine detection and fixed with 1% formaldehyde (Tousimis).

Transmitted/founder variant characterization. The number of transmitted/founder variants were characterized blindly, as previously described²⁸, and deposited in GenBank under accession numbers KJ201031 to KJ201503.

Determination of susceptibility of circulating SIV to IFN- α . Freshly isolated, CD8-depleted PBMCs from naive rhesus macaque donors were stimulated for 3 days with 5 μ g ml^{−1} PHA and 100 U ml^{−1} IL-2 in RPMI supplemented with 10% FBS, 2 mM L-glutamine, 100 U ml^{−1} penicillin and 100 μ g ml^{−1} streptomycin (RPMI-Complete). Target cells were re-suspended at 10⁶ cells ml^{−1} in RPMI-Complete containing various concentrations of recombinant human IFN- α (PBL Interferon Source) and incubated at 37 °C for 4 h. IFN- α -containing culture supernatants were collected and stored at 37 °C. Target cells were split into duplicate cultures and spinoculated for 2 h at 800g with diluted, equivalent input amounts of SIV from plasma from 14 to 18 d.p.i. from IFN- α 2a or placebo macaques. Cells were washed and re-suspended in matched, stored IFN- α containing supernatants supplemented with 100 U ml^{−1} IL-2. After incubation at 37 °C for 7 days, cell-free culture supernatants were collected and a SIV p27 antigen capture assay was used to detect the presence of viral p27 antigen according to the manufacturer's instructions (ABL).

Binding and neutralizing antibody assays. IFN-binding antibodies were assessed as previously described³¹. To evaluate for neutralizing antibodies, A549 cells (American Type Culture Collection, Manassas, VA) were seeded at 1×10^4 per well in RPMI with 2% FBS and 2 mM L-glutamine and incubated for 24 h. For the standard curve, IFN- α 2b (Hoffman La Roche, Nutley, NJ) was added to cells in twofold serial dilution from 5 IU ml^{−1} to 0.04 IU ml^{−1}. For measurement of IFN- α 2b-neutralizing antibodies in plasma, IFN- α 2b was added at a concentration of 0.5 IU ml^{−1} along with 160-fold diluted plasma. Control wells received medium only or 5 IU ml^{−1} IFN- α 2b and 13.1 μ g ml^{−1} of a control neutralizing antibody. After an additional 24 h, the media was removed and replaced with RPMI 1640 containing 2% FBS, 2 mM L-glutamine containing encephalomyocarditis virus (EMCV, American Type Culture Collection, Manassas, VA) at a multiplicity of infection of 0.5. Cells were stained with crystal violet at 52 h and assessed for cytopathic effect as measured by optical density at 570 nm.

Quantitative RT-PCR. RNA was extracted from PBMCs preserved in TRIzol (Life Technologies) or from thawed cryopreserved PBMCs by RNeasy RT (Molecular Research Center, Inc.) according to the manufacturers' instructions. Purified RNA was added directly to a one-step quantitative RT-PCR reaction containing iScript RT-iTaq Taq enzyme mix (BioRad). MX1, OAS2 and β 2 microglobulin were labelled with a 5' FAM reporter and 3' BHQ1 quencher (Biosearch Technologies). We used the following oligonucleotide sequences: MX1 F AGGAGTTGCCCTTCCCAGA, MX1 R CCTCTGAAGCATCCGAAATC, MX1 P TGACCAGATGCCCGCTGGT G; OAS2 F CAGTCTGGTGGTGTGAGT, OAS2 R CAGCGAGGGTAAATCC TTGA, OAS2 P GCACTGGCATCAACAGTGCCAGA.

MX1 and OAS2 forward and reverse primers were used at 500 nM, and probes at 200 nM. Samples were run on an Applied Biosystems Sequence Detection System 7900HT (ABI). Expression levels of MX1 and OAS2 were normalized to β 2 microglobulin and calculated based on the $\Delta\Delta$ CT method.

Transcriptome analysis. Total RNA was prepared as described above. Polyadenylated transcripts were purified on oligo-dT magnetic beads, fragmented, reverse transcribed using random hexamers and incorporated into barcoded cDNA libraries based on the Illumina TruSeq platform. Libraries were validated by microelectrophoresis, quantified, pooled and clustered on Illumina TruSeq v2 flowcells. Clustered flowcells were sequenced on an Illumina HiSeq 2000 in 100-base single-read reactions.

New rhesus macaque genome. RNA-Seq data were analysed by alignment to a provisional assembly (deposited under BioProject accession PRJNA214746) and annotation of a new Indian *Macaca mulatta* genome (data provided by R.B.N., University of Nebraska Medical Center and A. Zimin, University of Maryland). Comparison of

RNA-seq data generated using the v.4 assembly demonstrated a greater absolute number of mapping reads, higher proportion of mapped reads per sample than to the hg19 RefSeq or RhesMac2 assembly (Supplementary Fig. 1 and Supplementary Information).

RNA-seq data analysis. RNA-seq data were submitted to The Gene Expression Omnibus (GEO) repository at the National Center for Biotechnology Information (NCBI). RNA-seq data were aligned to a provisional assembly of Indian *Macaca mulatta* (MuSuRCA rhesus assembly v.4) using STAR version 2.3.0e³²; parameters were set using the annotation as a splice junction reference, un-annotated non-canonical splice junction mappings and non-unique mappings were removed from downstream analysis. Transcripts were annotated using the provisional UNMC annotation v4.12. Transcript assembly, abundance estimates and differential expression analysis were performed using Cufflinks v2.1.1 and Cuffdiff³³. Samples with <49% mapped reads or exhibiting considerable 3' or 5' bias were excluded from further analysis. To reduce normalization bias due to varying read depths, samples were analysed in two separate groups: group 1 comprised 166 samples and contained the samples from PBMCs from the placebo + SIV, IFN-1ant + SIV, and pegylated IFN- α 2a-treated + SIV animals, and from samples for the analysis of PBMCs, LN CD4 T cells and rectal biopsies of uninfected, IFN- α 2a-treated animals; the average number of mapped reads was 12,188,890 (range: 3,237,374–87,371,564). Group 2 comprised the lymph node CD4 T cells from the three SIV-infected groups consisting of 37 samples; the average mapped read count for this group was 4,898,448 (range: 1,173,455–19,310,483). Differentially expressed genes were defined by pair-wise comparison of each time point to the Day 0 baseline. We included genes that had any acute time point (0 vs 7, 10, 21, 28) that was significantly differentially expressed by a fixed discovery rate-corrected *P* value (*q* value) < 0.05. Differential gene lists were uploaded to Ingenuity Pathway Analysis software (v1.0 Ingenuity Systems, <http://www.ingenuity.com/>) and pathways with significant enrichment by Fisher's exact test and the Benjamini–Hochberg multiple testing correction were identified. Heat maps and other visualization were generated using Partek Genomics Suite v6.6.

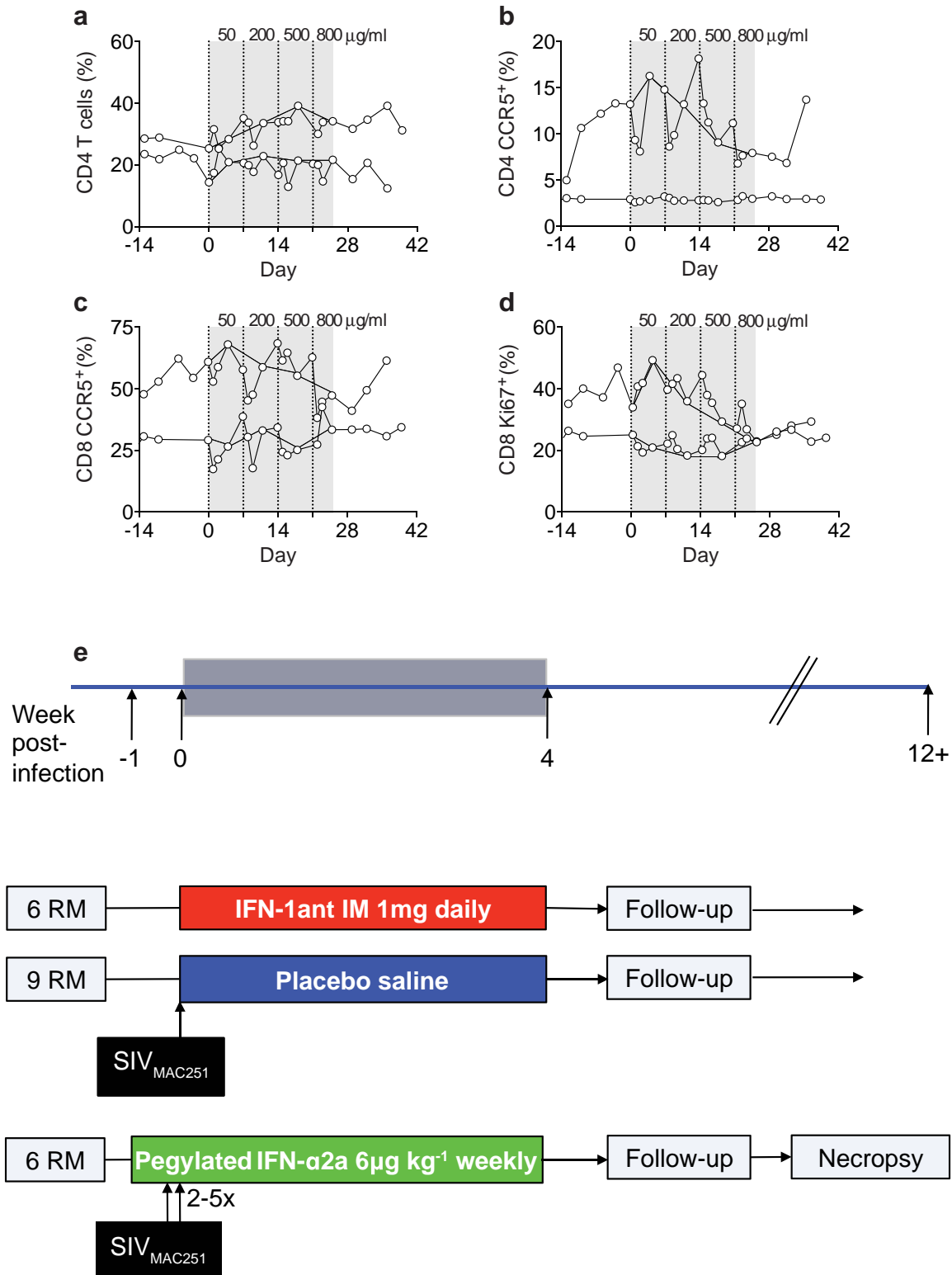
Gene-set enrichment analysis (GSEA). GSEA was performed using the desktop module available from the Broad Institute (<http://www.broadinstitute.org/gsea/>). FPKMs for samples from the IFN- α 2a and placebo groups at 7 d.p.i. were pre-filtered to remove transcripts with insufficient read coverage, and then were ranked using the signal-to-noise statistic. The gene-set was comprised of genes determined to be upregulated in unstimulated *FOXO3*^{−/−} macrophages relative to macrophages from control mice²⁵. Significance was estimated using gene-set permutation.

SIV *in situ* hybridization, immunohistochemistry and quantitative image analysis. SIV *in situ* hybridization was performed as previously described³⁴. Immunohistochemistry for rabbit polyclonal anti-APOBEC3G (Prestige Antibodies Powered by Atlas Antibodies HPA001812; Sigma-Aldrich), TRIM5 α (Prestige Antibodies Powered by Atlas Antibodies HPA023422; Sigma-Aldrich) and MX2 (Prestige Antibodies Powered by Atlas Antibodies HPA030235; Sigma-Aldrich) were performed using a biotin-free polymer approach (Rabbit Polink-2, Golden Bridge International, Inc.) on 5 μ m tissue sections mounted on glass slides, which were

dewaxed and rehydrated with double-distilled H₂O. Antigen retrieval was performed by heating sections in 0.01% citraconic anhydride containing 0.05% Tween-20 in a pressure cooker set at 122 °C for 30 s. Slides were rinsed in ddH₂O, incubated with blocking buffer (TBS containing 0.25% casein) and incubated with diluted rabbit anti-APOBEC3G, rabbit anti-TRIM5 α or rabbit anti-MX2 in blocking buffer overnight at 4 °C. Tissue sections were rinsed in wash buffer (1 \times TBS containing 0.05% Tween-20) for 10 min followed by an endogenous peroxidase blocking step using 1.5% (v/v) H₂O₂ in TBS (pH 7.4) for 10 min and placed in wash buffer. Slides were incubated with rabbit Polink-2 HRP polymer-staining system (Golden Bridge International, Inc.) according to manufacturer's recommendations (20–30 min at room temperature) then rinsed in wash buffer. Tissue sections were developed with Impact 3,3'-diaminobenzidine (Vector Laboratories), counterstained with haematoxylin and mounted in Permount (Fisher Scientific). All stained slides were scanned at high magnification (\times 200) using the ScanScope CS System (Aperio Technologies, Inc.) yielding high-resolution digital scans of the entire tissue section. Regions of interest of defined area (ROIs; 500 μ m²) were saved on the digital image using the Aperio rectangle tool (representing nearly the entire lymph node section) and high-resolution images were extracted from the ROIs of each whole-tissue scan. The per cent area of the lymph node (all anatomical compartments were included) that stained for APOBEC3G, TRIM5 α and MX2 were quantified under blind analysis using Photoshop CS5 and Fovea tools.

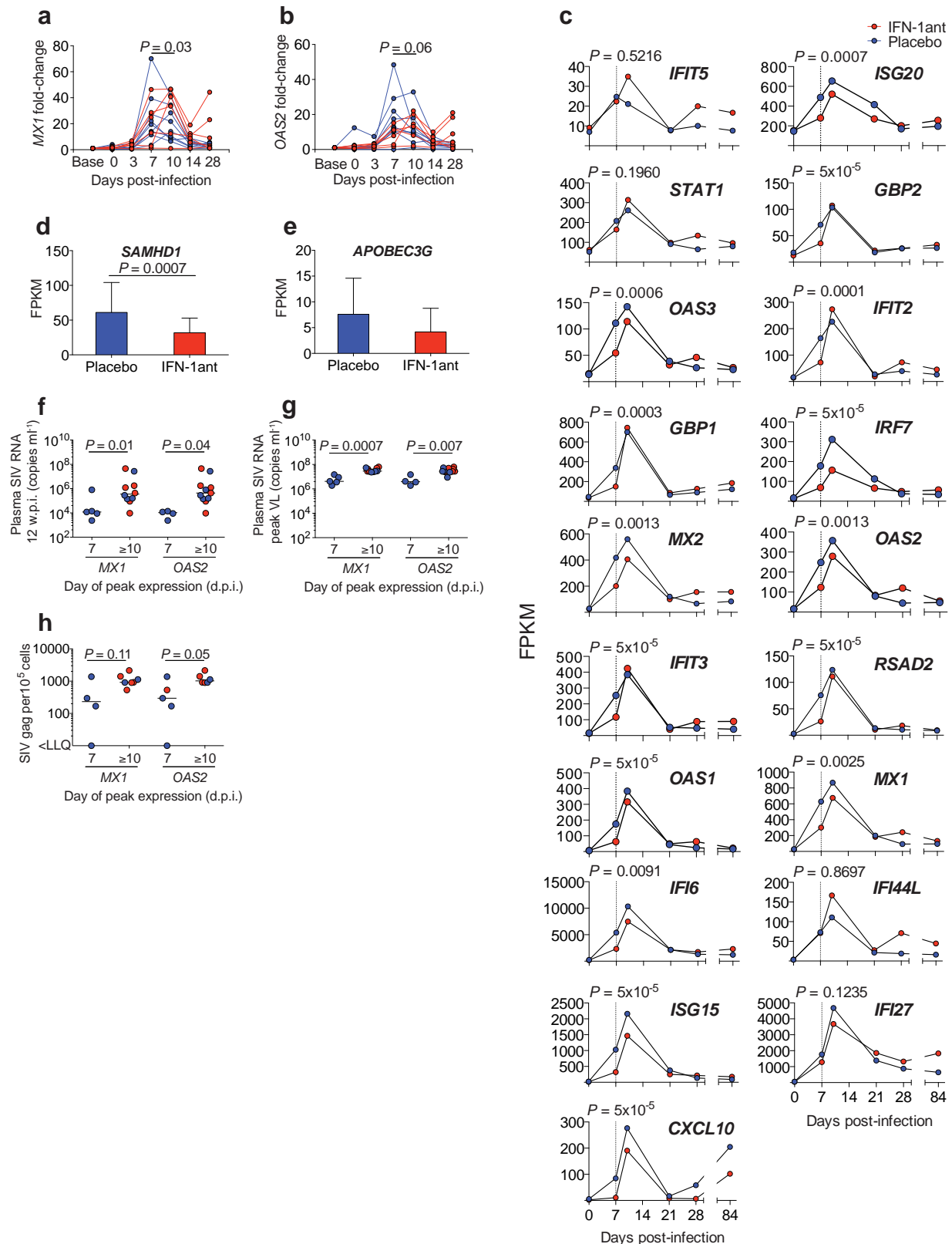
Statistical methods. Based on our previous data of rhesus macaques with acute SIV infection, the standard deviation for SIV RNA levels is 0.5×10^6 copies ml^{−1}. Using this value, 6 macaques in the IFN-1ant or IFN- α 2a group and 9 macaques in the placebo group would give us 80% power to detect a 0.8×10^6 copies ml^{−1} difference in SIV RNA levels between IFN-1ant or IFN- α 2a and placebo groups. Macaques were assigned to their respective groups randomly. Experiments, except as noted above, were not performed blindly. All replicates are biological replicates. Each experiment was performed once. Comparisons between groups at singular time points were performed with the Mann–Whitney *U* test, comparisons within groups with the Wilcoxon matched-pairs signed rank test, survival curve comparisons for per cent survival (IFN-1ant) and per cent uninfected (IFN- α 2a) with the log-rank (Mantel-Cox) test and correlations with Spearman coefficient, all using GraphPad Prism v5.0d. Comparisons of AUCs were performed using linear regression analysis adjusting for baseline values on JMP v10.

31. Vanderford, T. H. *et al.* Treatment of SIV-infected sooty mangabeys with a type-I IFN agonist results in decreased virus replication without inducing hyperimmune activation. *Blood* **119**, 5750–5757 (2012).
32. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
33. Trapnell, C. *et al.* Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nature Biotechnol.* **31**, 46–53 (2013).
34. Brenchley, J. M. *et al.* Differential infection patterns of CD4+ T cells and lymphoid tissue viral burden distinguish progressive and nonprogressive lentiviral infections. *Blood* **120**, 4172–4181 (2012).



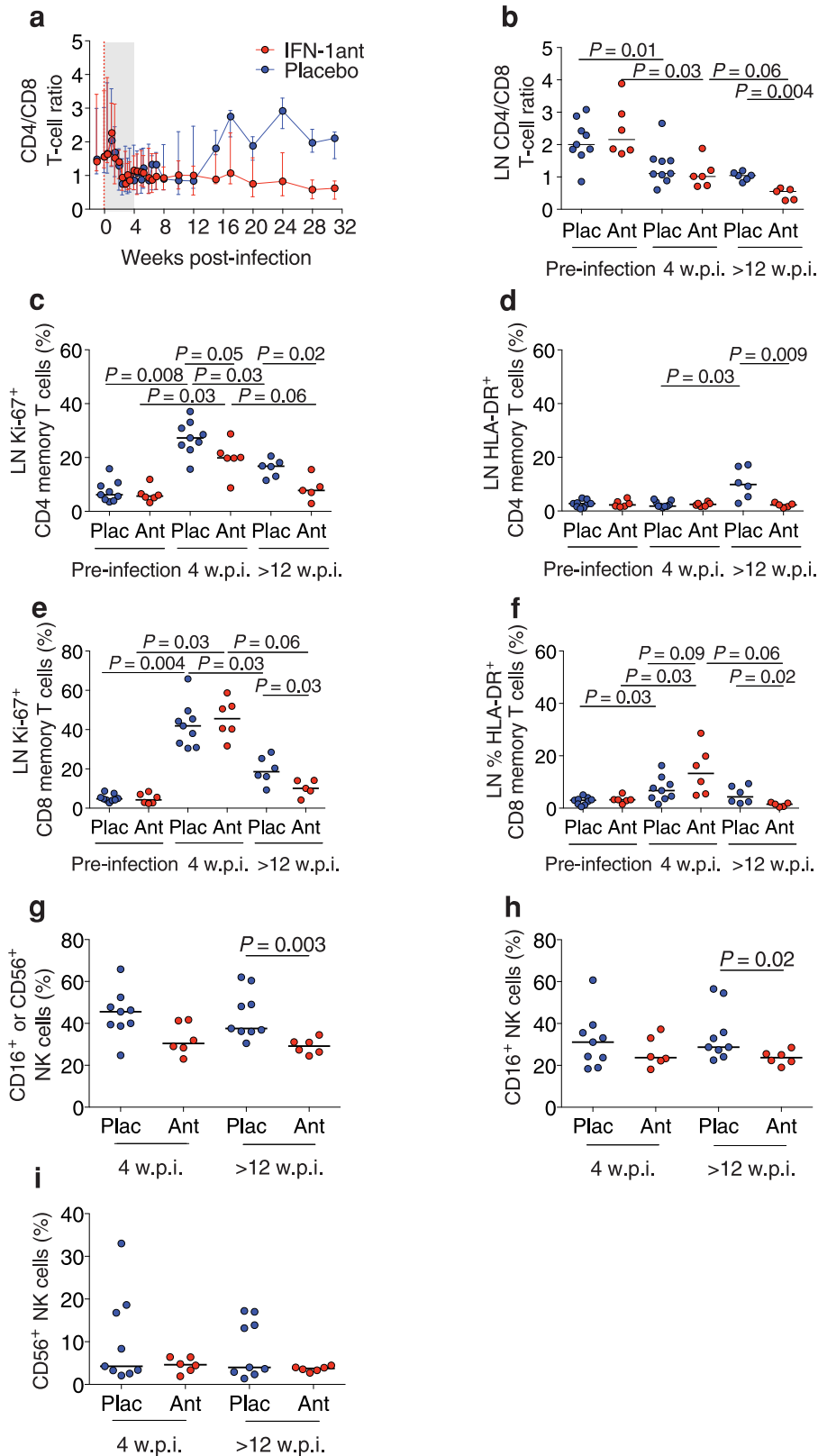
Extended Data Figure 1 | Dose escalation study for IFN-1ant and experimental schema. **a–d**, Effects of three times weekly IFN-1ant dosing on the frequency of CD4 T cells (**a**), CCR5⁺ CD4 T cells (**b**), CCR5⁺ CD8 T cells (**c**) and Ki67⁺ CD8 T cells (**d**) in 2 rhesus macaques. Dose was 50 µg in week 1, 200 µg in week 2, 500 µg in week 3 and 800 µg in week 4. Vertical dotted lines indicate the days a new dose was started. Black lines connect time points 4 days after the first dose. Grey shading indicates treatment period. **e**, Six macaques received 4 weeks of IFN-1ant intramuscularly starting at day 0 and were challenged intrarectally with 1 ml of a 1:25 dilution of SIV_{MAC251} (stock

concentration 3×10^8 SIV RNA copies ml⁻¹) at day 0 and followed until developing end-stage AIDS. Nine macaques were treated with 4 weeks of placebo saline intramuscularly starting at day 0 and challenged intrarectally with SIV_{MAC251} at day 0 and followed. Six macaques were injected weekly with IFN-α2a starting 1 week before the first challenge and through 4 w.p.i. Macaques required 2, 3 or 5 challenges to acquire systemic infection. Thus, macaques received 6, 7 or 9 doses of IFN-α2a. Macaques were necropsied at 12 w.p.i. per protocol.



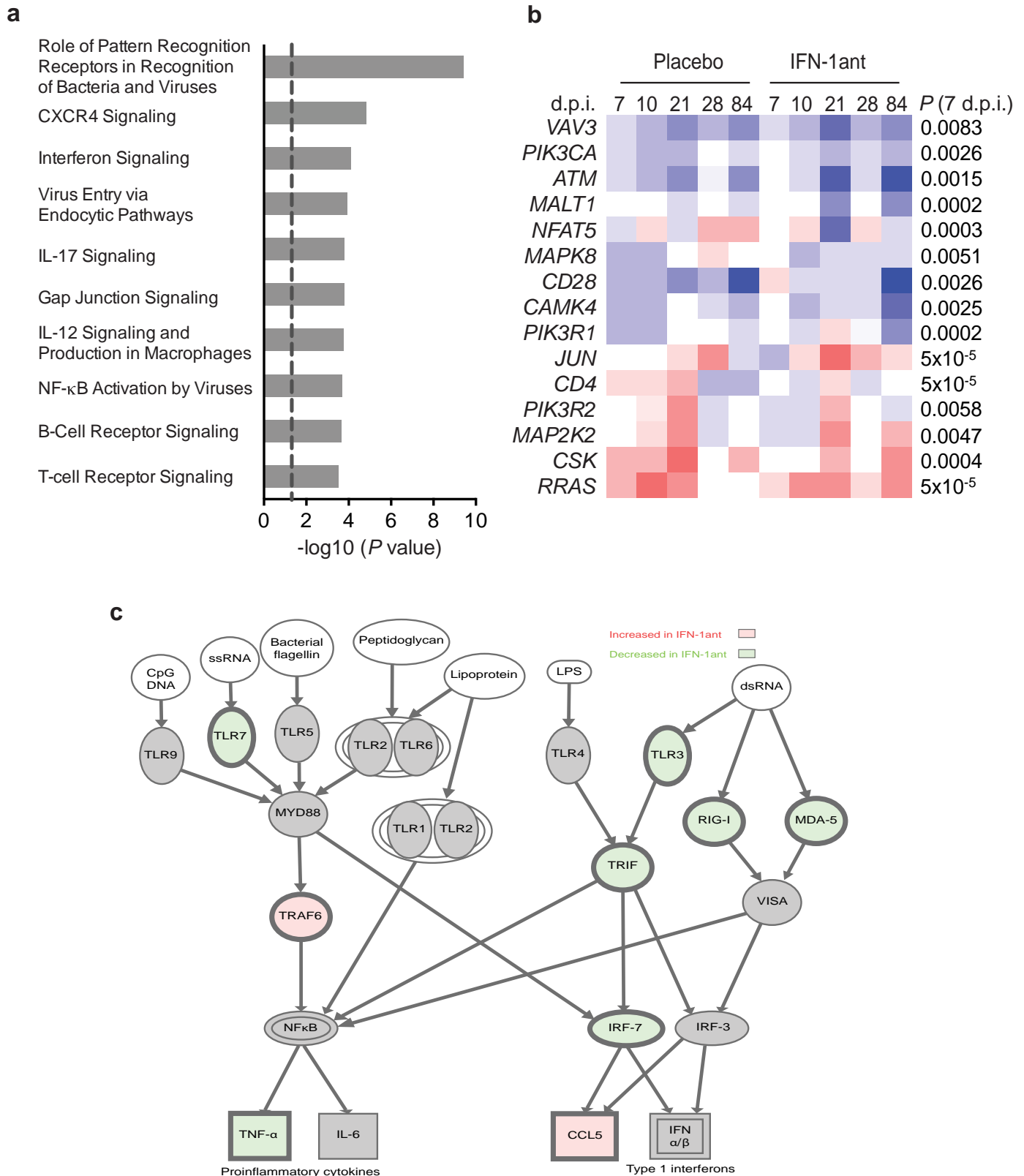
Extended Data Figure 2 | Effects of IFN-1ant on IFN-stimulated genes and virus burden. **a, b**, *MX1* (**a**) and *OAS2* (**b**) expression by qRT-PCR during acute SIV infection in IFN-1ant (red, $n = 6$) and placebo (blue, $n = 9$) macaques. P values were calculated by Mann-Whitney U test. **c**, ISGs in PBMCs in IFN-1ant and placebo macaques. P values represent the comparison between IFN-1ant ($n = 6$) and placebo ($n = 9$) macaque FPKMs at 7 d.p.i. **d, e**, *SAMHD1* (**d**) and *APOBEC3G* (**e**) expression in the lymph nodes in IFN-1ant ($n = 6$) and placebo ($n = 9$) macaques. P values were calculated by

Mann-Whitney U test. **f, g**, Plasma SIV RNA levels at 12 w.p.i. (**f**) or at peak (**g**) stratified by the day that *MX1* or *OAS2* expression peaked in PBMCs in IFN-1ant ($n = 6$) and placebo ($n = 9$) macaques. VL, viral load. P values were calculated by Mann-Whitney U test. **h**, SIV gag levels in PBMCs stratified by the day that *MX1* or *OAS2* expression peaked in PBMCs in IFN-1ant ($n = 6$) and placebo ($n = 6$) macaques. P values were calculated by Mann-Whitney U test. For all panels, IFN-1ant-treated macaques are represented in red, placebo-treated macaques in blue.



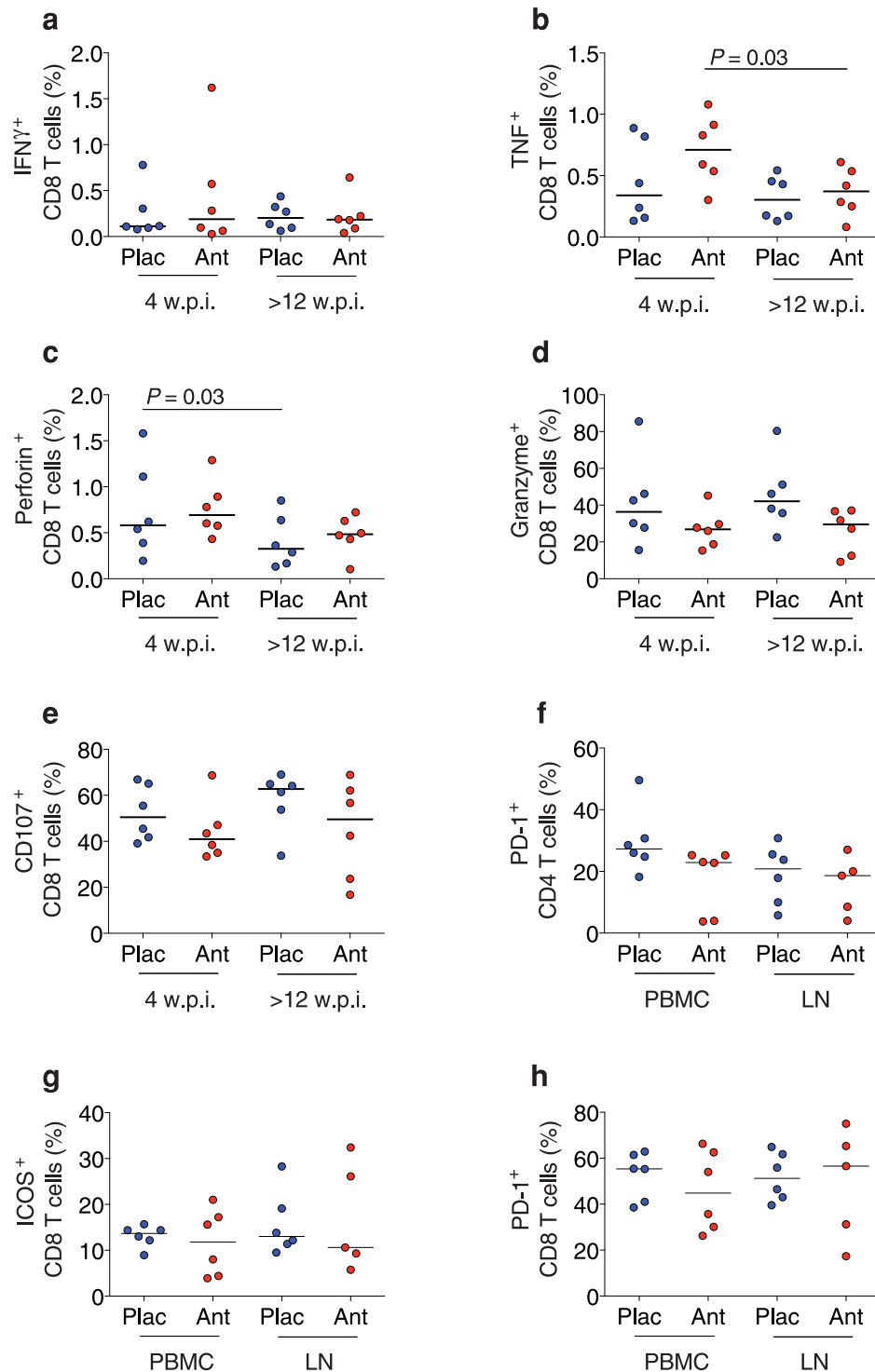
Extended Data Figure 3 | Effects of IFN-1ant on CD4 T cells and on immune activation. **a, b**, CD4/CD8 T-cell ratio in peripheral blood (**a**) and lymph node (LN) (**b**) in IFN-1ant (Ant, $n = 6$) and placebo (Plac, $n = 9$) macaques. Shading indicates treatment period. Error bars indicate range. Red vertical line indicates day 0 of systemic SIV infection. For all panels, horizontal bars indicate median values, and P values at different time points within treatment groups were calculated by Wilcoxon matched pairs signed rank test and between groups by Mann–Whitney U test. **c–f**, T-cell activation in lymph

nodes (**c–f**) in CD4 (**c, d**) and CD8 (**e, f**) T cells as represented by the frequency of Ki67⁺ (**c, e**) or HLA-DR⁺ (**d, f**) cells in IFN-1ant ($n = 6$) and placebo ($n = 9$) macaques. **g**, Frequency of circulating CD16⁺ or CD56⁺CD3[−]CD14[−] NK cells in IFN-1ant ($n = 6$) and placebo ($n = 9$) macaques. **h**, Frequency of circulating CD16⁺ NK cells in IFN-1ant ($n = 6$) and placebo ($n = 9$) macaques. **i**, Frequency of circulating CD56⁺ NK cells in IFN-1ant ($n = 6$) and placebo ($n = 9$) macaques. For all panels, IFN-1ant-treated macaques are represented in red, placebo-treated macaques in blue.



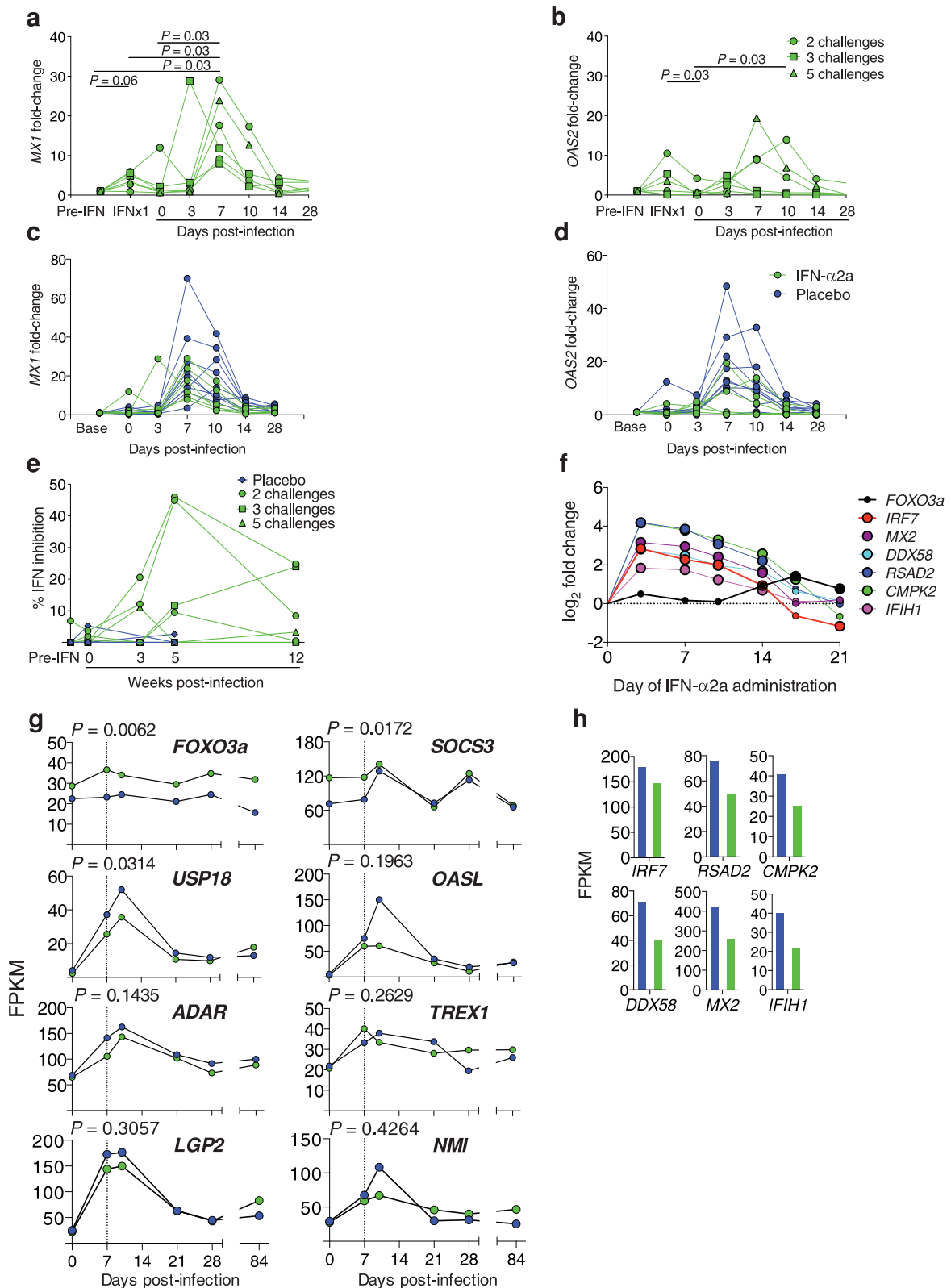
Extended Data Figure 4 | IFN-1ant alters innate and adaptive immune signalling. **a**, Selected pathways significantly affected by IFN-I blockade. P values were calculated by Fisher's exact test with the Benjamini–Hochberg multiple testing correction. **b**, Expression of genes involved in pattern recognition receptor signalling of IFN-1ant-treated macaques ($n = 6$)

compared to placebo ($n = 9$) at 7 d.p.i. Upregulation compared to pre-infection is represented by red, no change by white, downregulation by blue. P values represent the comparison between IFN-1ant and placebo macaques at 7 d.p.i. **c**, Selected genes in pattern recognition receptor signalling pathways. Upregulation at 7 d.p.i. is represented by red, downregulation by green.



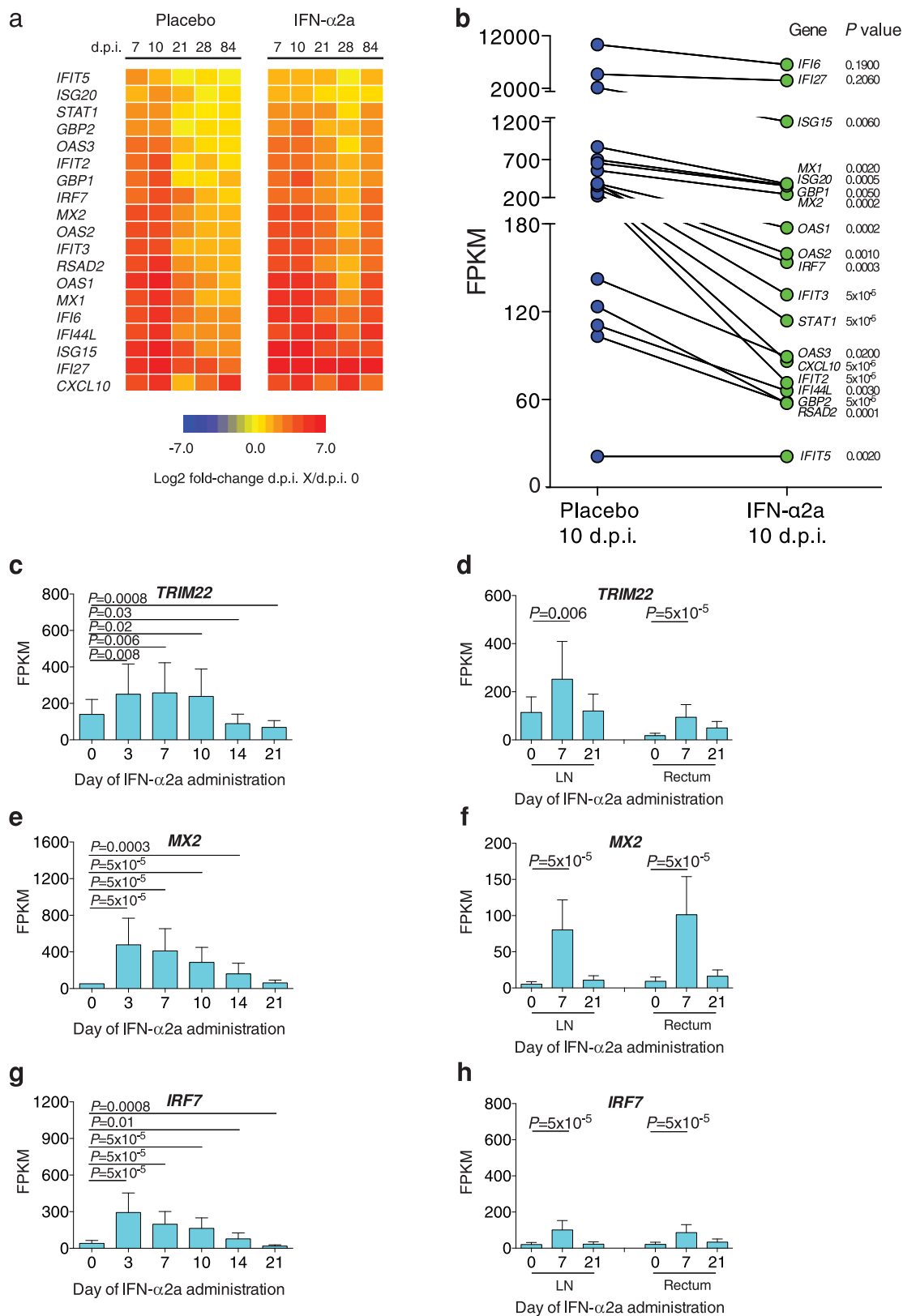
Extended Data Figure 5 | Effects of IFN-1ant on T-cell function and phenotype. **a–e**, SIV-specific responses in peripheral blood at 4 and >12 w.p.i. in IFN-1ant (Ant, $n = 6$) and placebo (Plac, $n = 6$) macaques by frequency of IFN- γ^+ (**a**), TNF $^+$ (**b**), perforin $^+$ (**c**), granzyme B $^+$ (**d**) and CD107 $^+$ (**e**) CD8 T cells. T-cell exhaustion in peripheral blood and lymph nodes (LN) at >16 w.p.i.

based on frequency of PD-1 $^+$ CD4 (**f**) and CD8 (**h**) T cells and ICOS $^+$ (**g**) CD8 T cells. For all panels, P values at different time points within treatment groups were calculated by Wilcoxon matched pairs signed rank test and between groups by Mann-Whitney U test. IFN-1ant-treated macaques are represented in red, placebo-treated macaques in blue.



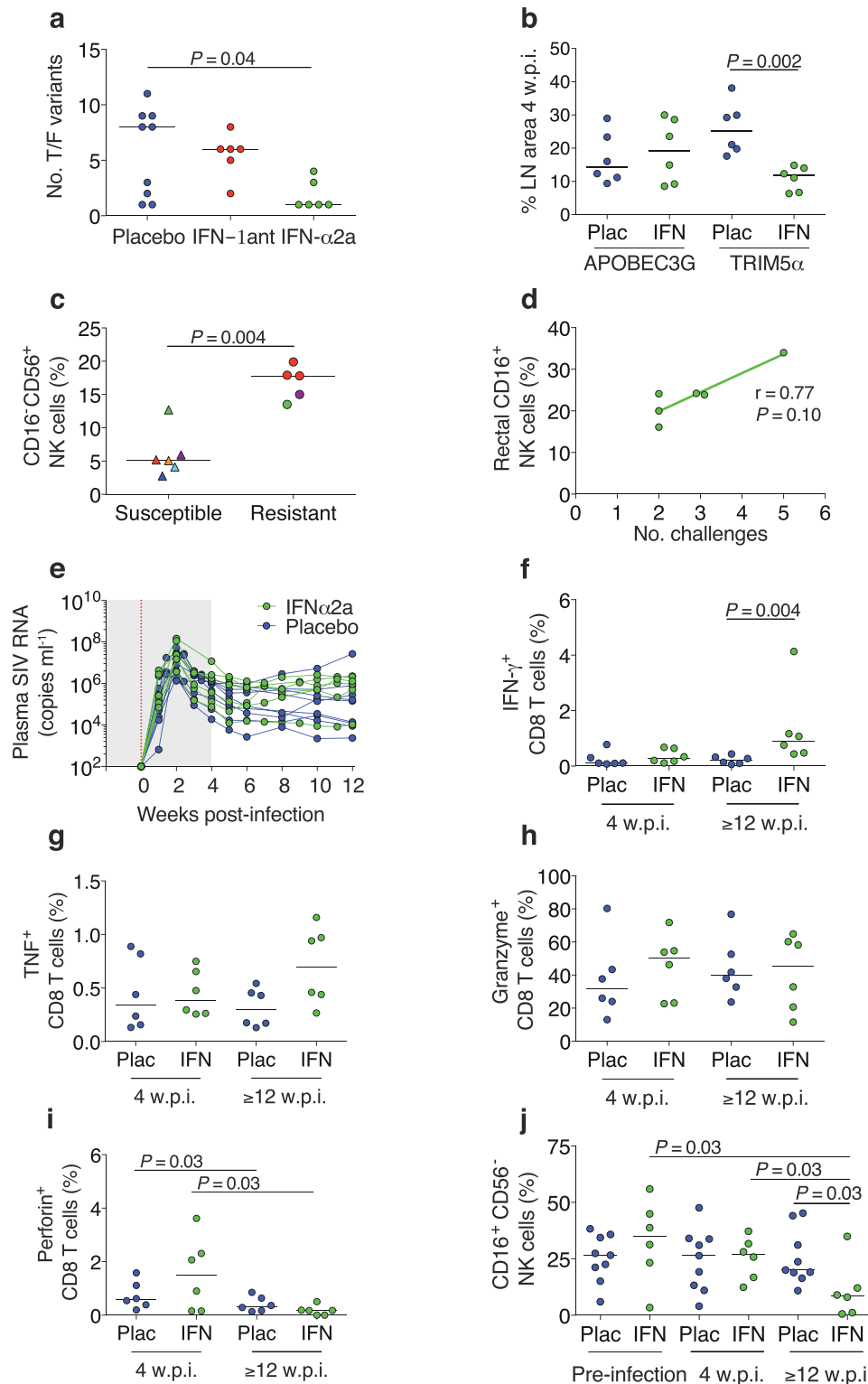
Extended Data Figure 6 | IFN- α 2a treatment transiently induces ISGs and subsequently induces the IFN-repressor FOXO3a but does not induce neutralizing anti-IFN antibodies. a–d, MX1 (a, c) and OAS2 (b, d) expression during the duration of IFN- α 2a treatment in the IFN- α 2a group alone (a, b) and during infection in the IFN- α 2a ($n = 6$) and placebo ($n = 9$) groups (c, d). P-values were calculated by Wilcoxon matched pairs signed rank test. e, Percentage of *in vitro* IFN antiviral activity inhibited by plasma from IFN- α 2a ($n = 6$) and placebo ($n = 3$) macaques. f, Expression of FOXO3a and

FOXO3a-bound genes in SIV-uninfected macaques ($n = 3$) treated with 21 days of IFN- α 2a. Large circles indicate statistically significant ($P < 0.05$) changes from pre-IFN- α 2a treatment calculated by Wilcoxon matched pairs signed rank test. Small circles indicate no statistically significant change from pre-IFN- α 2a treatment. g, Expression of IFN- α -regulatory genes in IFN- α 2a ($n = 6$) and placebo ($n = 9$) macaques. P-values represent the comparison between FPKMs of IFN- α 2a ($n = 6$) and placebo ($n = 9$) macaques at 7 d.p.i. h, Expression of FOXO3a-bound genes in IFN- α 2a (green, $n = 6$) and placebo (blue, $n = 9$) macaques at 7 d.p.i.



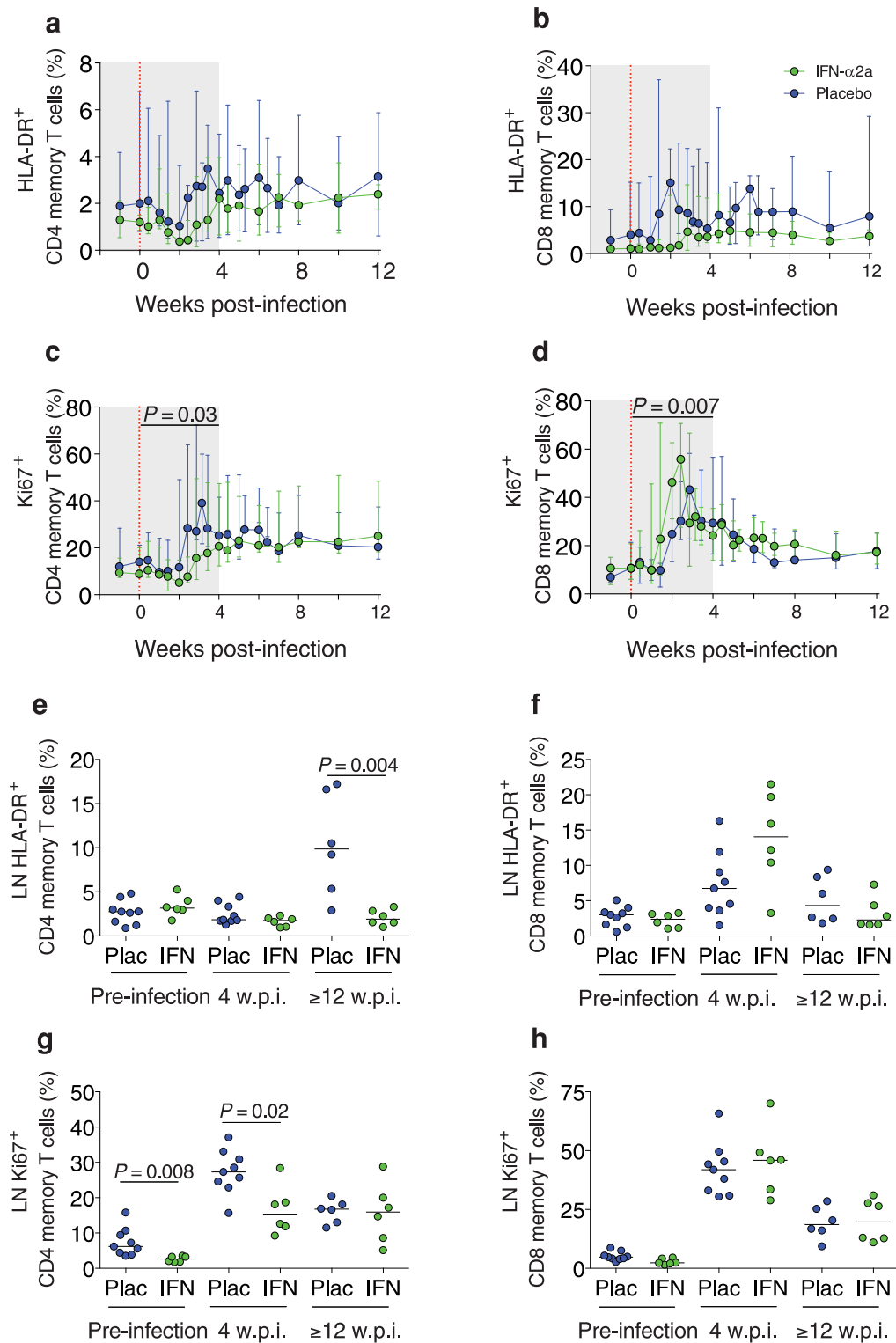
Extended Data Figure 7 | Effects of IFN- α 2a on IFN-stimulated and antiviral genes. **a**, ISGs in PBMCs in IFN- α 2a ($n = 6$) and placebo ($n = 9$) macaques. Red indicates upregulation, yellow indicates no change and blue indicates downregulation relative to pre-infection. **b**, Expression of ISGs in macaques treated with IFN- α 2a ($n = 6$) or placebo ($n = 9$). P values indicate differentially expressed genes at 10 d.p.i. **c–h**, Expression of *TRIM22*

(**c**, **d**), *MX2* (**e**, **f**) and *IRF7* (**g**, **h**) in SIV-uninfected macaques ($n = 3$) treated with weekly IFN- α 2a for 3 weeks in PBMCs (**c**, **e**, **g**) and lymph nodes and rectum (**d**, **f**, **h**). Day 0 reflects baseline. Numbers indicate days since first IFN- α 2a administration. Error bars indicate range. P values were calculated by Wilcoxon matched pairs signed rank test.



Extended Data Figure 8 | Effects of IFN-α2a on SIV control. **a**, Number of transmitted/founder (T/F) variants in placebo ($n=9$), IFN-1ant ($n=6$) and IFN-α2a ($n=6$) macaques. P value was calculated by Mann-Whitney U test. **b**, Antiviral protein production in lymph nodes (LN) by immunohistochemistry at 4 w.p.i. in IFN-α2a ($n=6$) and placebo ($n=6$) macaques. P value was calculated by Mann-Whitney U test. **c**, CD56⁺ NK-cell frequency on the day of challenge stratified by whether the macaque resisted or was susceptible to systemic infection that day. Each IFN-α2a macaque ($n=6$) is indicated by a different colour. Circles indicate that the macaque was resistant to infection with the next challenge and triangles indicate that the macaque was susceptible to infection with the next challenge. P value was calculated by Mann-Whitney U test. **d**, Correlation between the number of

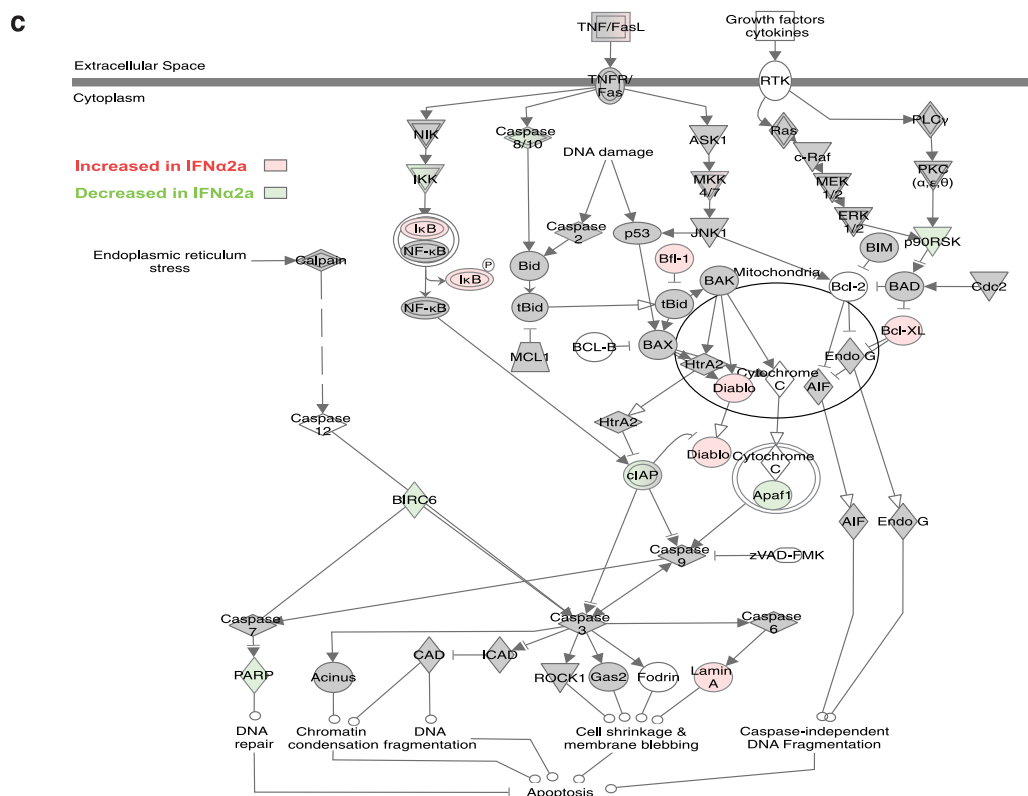
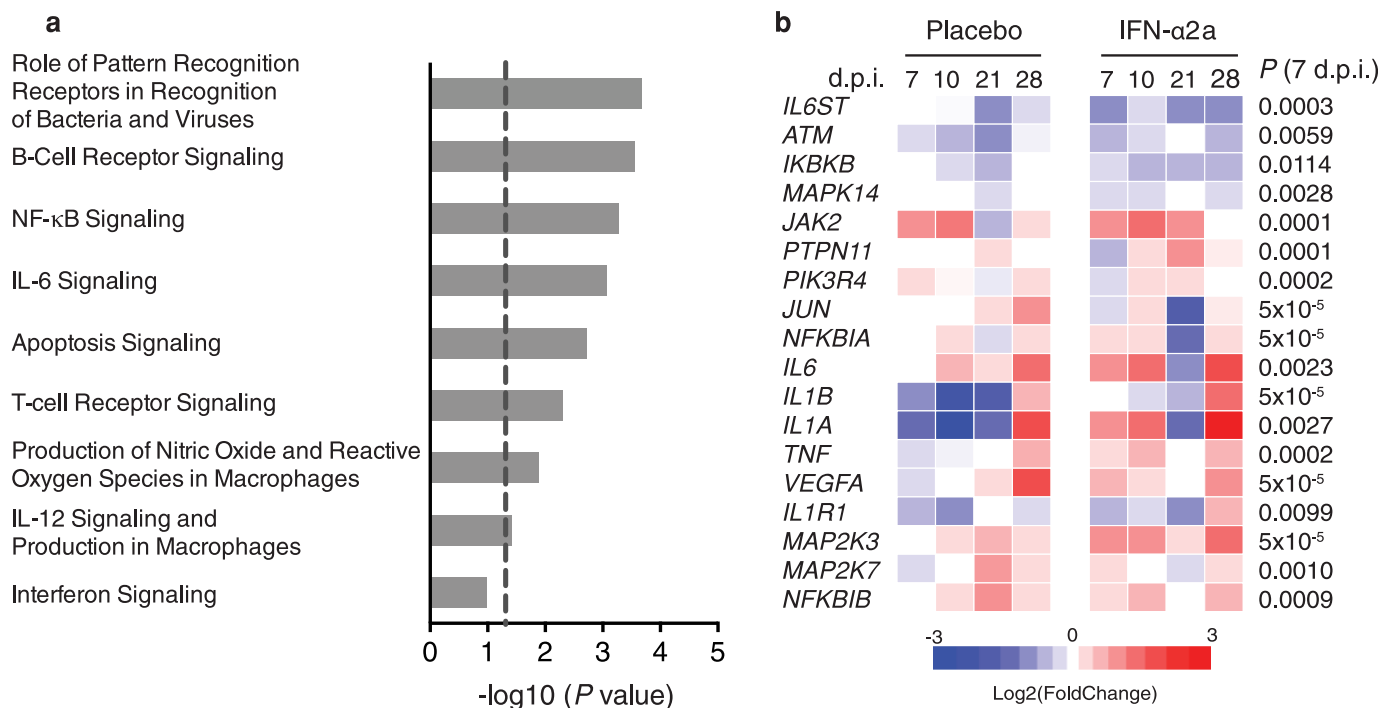
challenges required to achieve systemic infection and rectal CD16⁺ NK-cell frequency in each macaque ($n=6$) at 4 w.p.i. r indicates the Spearman's rank correlation coefficient. P value indicates the significance of the correlation. **e**, Plasma SIV RNA levels in macaques treated with IFN-α2a ($n=6$) or placebo ($n=9$) saline. Shading reflects treatment period. Red vertical line indicates day 0 of systemic SIV infection. **f–i**, Frequency of IFN-γ⁺ (**f**), TNF⁺ (**g**), granzyme B⁺ (**h**) and perforin⁺ (**i**) CD8 T cells at 4 and ≥12 w.p.i. in IFN-α2a ($n=6$) and placebo ($n=6$) macaques. **j**, Frequency of circulating CD16⁺CD56⁻ NK cells in IFN-α2a ($n=6$) and placebo ($n=9$) macaques. P values at different time points within treatment groups were calculated by Wilcoxon matched pairs signed rank test and between groups by Mann-Whitney U test.



Extended Data Figure 9 | Effects of IFN-α2a on T-cell activation.

a–h, Frequency of peripheral blood (**a–d**) and lymph node (LN) (**e–h**) CD4 (**a, c, e, g**) and CD8 (**b, d, f, h**) memory T cells expressing HLA-DR (**a, b, e, f**) or Ki67 (**c, d, g, h**) in IFN-α2a (IFN, $n = 6$) and placebo (Plac, $n = 9$) macaques.

Shading indicates treatment period. Error bars indicate range. **a–d**, Red vertical line indicates day 0 of systemic SIV infection. P values represent the comparison between groups of the AUC (0–4 w.p.i.). **e–h**, Horizontal bars indicate median values. P values were calculated by Mann–Whitney U test.



Extended Data Figure 10 | Effects of IFN- α 2a on gene expression. **a**, Selected pathways significantly affected by IFN- α 2a treatment. *P* values were calculated by Fisher's exact test with the Benjamini–Hochberg multiple testing correction. **b**, Expression of genes downstream of IL-6 signalling. Upregulation relative to before IFN- α 2a or placebo treatment and SIV infection is

represented by red, no change by white, downregulation by blue. *P* values represent the comparison between IFN- $\alpha 2a$ ($n = 6$) and placebo ($n = 9$) macaques at 7 d.p.i. **c**, Selected genes in apoptosis signalling pathways. Significant upregulation at 7 d.p.i. is represented by red, downregulation by green.

The DNA methylation landscape of human early embryos

Hongshan Guo^{1*}, Ping Zhu^{1,2*}, Liying Yan^{1,3*}, Rong Li^{1,3*}, Boqiang Hu¹, Ying Lian^{1,3}, Jie Yan^{1,3}, Xiulian Ren^{1,3}, Shengli Lin^{1,3}, Junsheng Li^{1,3}, Xiaohu Jin^{1,3}, Xiaodan Shi^{1,3}, Ping Liu^{1,3}, Xiaoye Wang⁴, Wei Wang⁴, Yuan Wei⁴, Xianlong Li¹, Fan Guo¹, Xinglong Wu¹, Xiaoying Fan¹, Jun Yong^{1,5}, Lu Wen¹, Sunney X. Xie^{1,5}, Fuchou Tang^{1,6} & Jie Qiao^{1,3}

DNA methylation is a crucial element in the epigenetic regulation of mammalian embryonic development^{1–5}. However, its dynamic patterns have not been analysed at the genome scale in human pre-implantation embryos due to technical difficulties and the scarcity of required materials. Here we systematically profile the methylome of human early embryos from the zygotic stage through to post-implantation by reduced representation bisulphite sequencing and whole-genome bisulphite sequencing. We show that the major wave of genome-wide demethylation is complete at the 2-cell stage, contrary to previous observations in mice. Moreover, the demethylation of the paternal genome is much faster than that of the maternal genome, and by the end of the zygotic stage the genome-wide methylation level in male pronuclei is already lower than that in female pronuclei. The inverse correlation between promoter methylation and gene expression gradually strengthens during early embryonic development, reaching its peak at the post-implantation stage. Furthermore, we show that active genes, with the trimethylation of histone H3 at lysine 4 (H3K4me3) mark at the promoter regions in pluripotent human embryonic stem cells, are essentially devoid of DNA methylation in both mature gametes and throughout pre-implantation development. Finally, we also show that long interspersed nuclear elements or short interspersed nuclear elements that are evolutionarily young are demethylated to a milder extent compared to older elements in the same family and have higher abundance of transcripts, indicating that early embryos tend to retain higher residual methylation at the evolutionarily younger and more active transposable elements. Our work provides insights into the critical features of the methylome of human early embryos, as well as its functional relation to the regulation of gene expression and the repression of transposable elements.

DNA methylation is an important form of epigenetic modification and has a crucial role in many biological processes, including repression of gene transcription, maintenance of gene imprinting and X-chromosome inactivation, and repression of transposable elements^{1–7}. The most dramatic genome-wide changes of the methylome in mammals occur in primordial germ cells and during pre-implantation development^{3,8–13}.

To obtain DNA methylation maps of human early embryos, we performed reduced representation bisulphite sequencing (RRBS) on human gametes, as well as pre- and post-implantation embryos (Supplementary Tables 1 and 2). The methylation profiles form four distinct clusters during early embryonic development (Extended Data Fig. 1a, b): the hypomethylated oocytes and polar bodies, the hypermethylated sperm, the hypomethylated cleavage-stage embryos, and the hypermethylated post-implantation embryos, which are similar to those found in mice⁹. Moreover, the overall DNA methylation level of the gene body (the genomic region from transcription start site (TSS) of a gene to its transcription end site (TES)) is higher than that of neighbouring intergenic regions and there is a markedly hypomethylated region around the TSS, similar

to patterns observed in other types of cells⁶. The methylation level of the gene body is relatively even, with a slight increase from the TSS to the TES and a clear reduction after the TES (Fig. 1a). Recently, it has been shown that mouse oocytes exhibit relatively high levels of non-CpG methylation in their genomes^{10,14}, but this has not been investigated in human oocytes. We found that in human mature oocytes, as well as first and second polar bodies, there are significant levels of non-CpG methylation around gene body regions in the genome; the enrichment pattern of this non-CpG methylation was similar to that of CpG methylation (Extended Data Fig. 1c, d). Moreover, the level of non-CpG methylation around gene bodies in the oocytes is clearly correlated with the level of expression of corresponding genes, indicating the potential functional significance of non-CpG methylation (Extended Data Fig. 1e).

Previous observations in mice show that the most marked demethylation occurs at the zygotic stage, with mild gradual demethylation from this point onwards until the blastocyst stage⁹. We found that the methylomes of human embryos are similar to those of mouse embryos, but do have distinct features. A dramatic decrease in DNA methylation occurs between fertilization and the 2-cell stage, with the average level of methylation decreasing from 54% in the sperm and 48% in metaphase II (MII) oocytes, to 41% in the zygotes and further to 32% in the 2-cell embryos (Fig. 1b and Extended Data Fig. 2a). Notably, there are subtle changes in the level of DNA methylation between the 2-cell and morula stages, which differs from findings in mice. A further reduction of DNA methylation to around 29% occurs as the embryo progresses from the morula stage to the blastocyst stage in the inner cell mass (ICM). Following implantation, a sharp increase in the level of methylation is observed (Fig. 1b and Extended Data Fig. 2b–d). The genomic regions with high CpG density tend to be hypomethylated, whereas those with low CpG density tend to be hypermethylated (Extended Data Fig. 2e–g). Notably, the genome-wide demethylation in mouse embryos occurs mainly at the 1-cell stage, whereas in human embryo the demethylation occurs from fertilization to the 2-cell stage. When we analysed the genomic regions, such as the genic and intergenic regions separately, the patterns of demethylation and re-methylation are similar to those found when analysing the genome-wide levels of demethylation, indicating that the dynamic changes in DNA methylation are in general universal throughout the entire genome (Extended Data Fig. 3a, b). Although RRBS analysis covers the majority of CpG islands (CGIs), which are probably the most informative methylation sites, it only covers around 10% of all the CpG sites in the human genome, leaving CpG-sparse regions unexplored. To resolve this issue and to obtain absolute quantification of DNA methylation at the whole-genome scale, we performed whole-genome bisulphite sequencing (WGBS) on the ICM and post-implantation embryos, which account for the lowest and highest methylation status of the genome during early embryonic development, respectively (Supplementary Table 3). We found that the genome-wide methylation levels in the ICM and post-implantation

¹Biodynamic Optical Imaging Center & Center for Reproductive Medicine, College of Life Sciences, Third Hospital, Peking University, Beijing 100871, China. ²Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China. ³Key Laboratory of Assisted Reproduction, Ministry of Education, Beijing 100191, China. ⁴Department of Obstetrics and Gynecology, Peking University Third Hospital, Beijing 100191, China. ⁵Department of Chemistry and Chemical Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ⁶Ministry of Education Key Laboratory of Cell Proliferation and Differentiation, Beijing 100871, China.

*These authors contributed equally to this work.

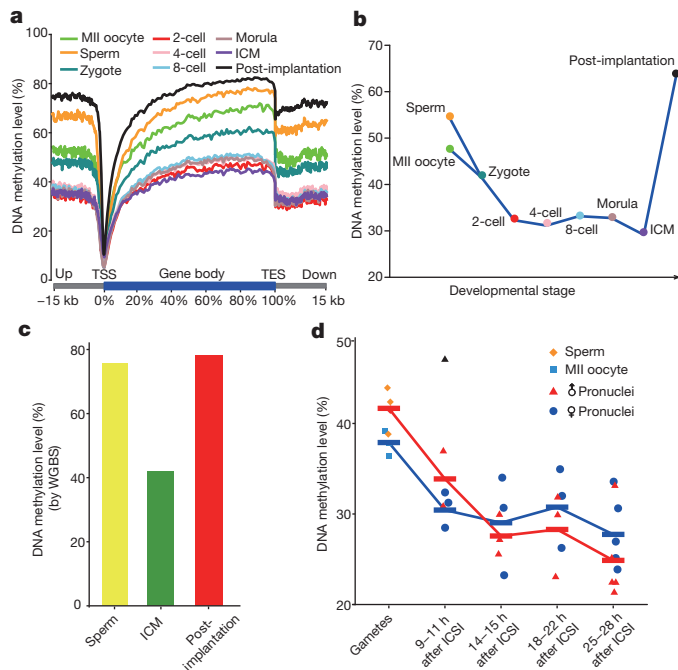


Figure 1 | Dynamics of the DNA methylome in human early embryos.

a, Averaged DNA methylation levels along the gene bodies and 15 kilobases (kb) upstream of the transcription start sites (TSS) and 15 kb downstream of the transcription end sites (TES) of all RefSeq genes. **b**, Methylation landscape across each stage of human early embryos. The averaged DNA methylation level of each developmental stage is calculated based on the overlapped 100-base-pair (bp) tiles detected in all of the developmental stages analysed. **c**, Averaged DNA methylation levels of human sperm and early embryos analysed by whole genome bisulphite sequencing (WGBS), including sperm (yellow bar; previously published²⁶), ICM of the blastocysts (green bar), and post-implantation embryos (liver, red bar). **d**, Averaged DNA methylation levels of individual male and female pronuclei of zygotes at different time points after intra-cytoplasmic sperm injection (ICSI). Note that the DNA methylation levels decrease dramatically in both male (red line) and female (blue line) pronuclei and that at the late pronuclear stage the genome-wide methylation levels of male pronuclei are already lower than those of female pronuclei. The diamond and square represent sperm and MII oocytes, while the triangle and circle represent male and female pronuclei, respectively. The black triangle represents one outlier of the male pronuclei. The averaged DNA methylation level of each individual pronucleus is calculated based on individual CpG sites covered by three or more sequencing reads.

embryos were 42.0% and 78.1%, respectively (Fig. 1c and Extended Data Fig. 4a), and that the distribution of DNA methylation on and around gene bodies accurately matched that inferred by RRBS (Fig. 1a and Extended Data Fig. 4b).

We also analysed DNA methylation patterns in the first and second polar bodies and found that they are comparable to those found in MII oocytes (Extended Data Fig. 4c). This indicates that during the extrusion of the first and second polar bodies from the oocytes, there is no genome-wide asymmetric pattern of DNA methylation between them. Similarly, the methylation levels of different genomic regions in the trophectoderm of the blastocysts are only slightly lower than those in the ICM (Extended Data Fig. 4d).

To gain further insight into the mechanism of the genome-wide demethylation, we analysed the methylomes of individual male and female pronuclei separately using a single-cell RRBS technique that we have recently developed¹⁵ (Extended Data Fig. 5a, b and Supplementary Table 3). We found that the demethylation process is very heterogeneous as the methylation levels of individual male (or female) pronuclei at the same time point following intra-cytoplasmic sperm injection show marked variation (Fig. 1d). Furthermore, the demethylation of the paternal genome is much faster than that of the maternal genome, and in late pronuclear

stage zygotes the genome-wide methylation level in male pronuclei is already lower than that in female pronuclei (Fig. 1d and Extended Data Fig. 5c). This was further verified by immunostaining (Extended Data Fig. 6). Interestingly, both male and female pronuclei showed some evidence of hydroxymethylation, although usually the signals in male pronuclei were stronger than those in female ones.

Next we analysed the similarities and differences in DNA methylation between sperm and oocytes. Sperm and oocytes exhibited comparable methylation patterns for 64.3% of all the covered 100-bp tiles, which are either hypermethylated (methylation levels $\geq 75\%$) or hypomethylated (methylation levels $\leq 25\%$) in both gametes (Extended Data Fig. 7a, b). Notably, the hypermethylated regions in both gametes exhibited similar DNA methylation dynamics as the mean genomic changes. These regions are strongly enriched in transposable elements such as short interspersed nuclear elements (SINEs) and long interspersed nuclear elements (LINEs) (Extended Data Fig. 7c), indicating that the hypermethylated regions in gametes of both sexes are mainly enriched in transposable elements, probably repressing their transcription and activity, as well as in introns, probably regulating gene transcription and splicing. In contrast, the hypomethylated regions are enriched in high-density CpG promoters,

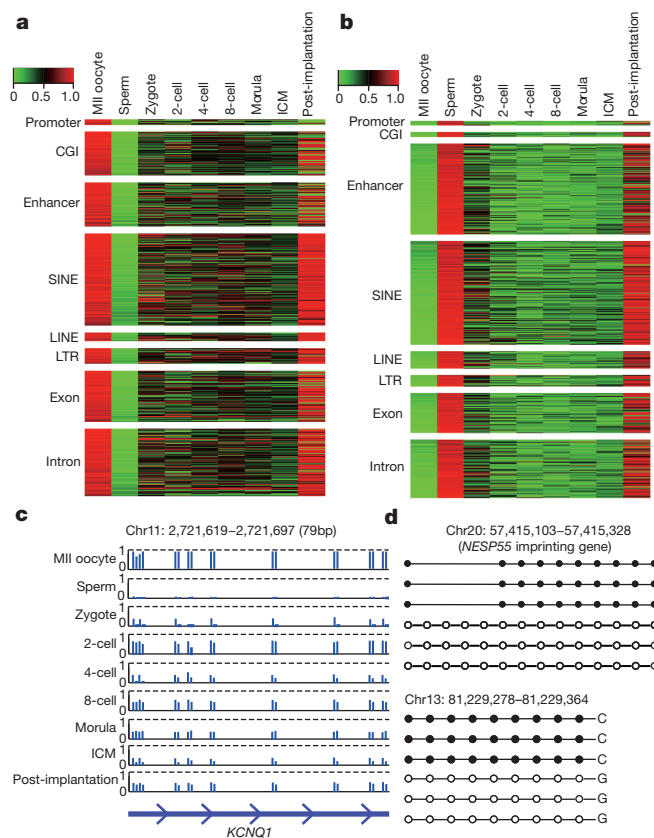


Figure 2 | Key features of gamete-specific differentially methylated regions (DMRs). **a**, Heat map of the methylation level of oocyte-specific DMRs among different genomic regions across different developmental stages. **b**, Heat map of the methylation level of sperm-specific DMRs among different genomic regions across different developmental stages. In panels **a** and **b**, the colour keys from green to red indicate low to high methylation level, respectively. **c**, Representative locus of a known maternal imprinting gene, *KCNQ1*, covered in our RRBS data set. The blue bars indicate the DNA methylation levels of different CpG sites. The region was fully methylated in MII oocytes, unmethylated in sperm cells and around 50% methylated in cleavage-stage embryos and post-implantation embryos. **d**, One known imprinting locus and one potential novel imprinting locus showing ASM, tracked with SNPs to distinguish their allele origins. The paired reads generated from the WGBS data sets with heterozygous SNPs were selected to show the DNA methylation levels of the two alleles separately.

enhancers of a wide variety of human tissues, exons, and CGIs, and are kept hypomethylated throughout early embryonic development.

We also examined differentially methylated regions (DMRs) between sperm and oocytes. In total, we identified 17,473 sperm-specific DMRs and 12,145 oocyte-specific DMRs (Extended Data Fig. 7d, e). The sperm-specific DMRs were relatively enriched in intergenic regions whereas oocyte-specific DMRs were enriched in intragenic regions (Fig. 2a, b and Extended Data Fig. 7f). Sperm-specific DMRs were strongly enriched for the tissue-specific enhancers marked by H3K4me1 in a wide variety of tissues compared with oocyte-specific DMRs. This indicates that sperm tends to hypermethylate enhancer elements that are active or permissive in tissues other than male germ cells, probably to avoid their aberrant activation during male germ cell development. As expected, oocyte-specific DMRs localize to CGIs more frequently than sperm-specific DMRs. Similar to those in mice, some of the oocyte-specific DMRs probably retain allele-specific DNA methylation (ASM) patterns during pre-implantation development with an average methylation level around 50%. On the contrary, the sperm-specific DMRs rapidly lose methylation before the 2-cell stage and retain only background levels of methylation. However, the

majority of both oocyte-specific and sperm-specific DMRs were re-methylated after implantation (Fig. 2a, b and Extended Data Fig. 7d, e).

Although RRBS analysis shows relatively poor coverage for imprinting genes, for the known imprinting DMRs it covers we found that the methylation levels are accurately maintained at around 50% from zygotic stage to post-implantation stage, as expected (Fig. 2c, d and Extended Data Fig. 8a, b). To extend the analyses to potentially novel imprinting genes, we took advantage of our WGBS data set from a post-implantation embryo and performed ASM analyses based on identified heterozygous SNPs. We found that known imprinted genes with enough sequencing coverage and available heterozygous SNPs showed ASM as expected. Moreover, we found 120 novel ASM regions (Fig. 2d, Extended Data Fig. 8c, d and Supplementary Table 4), some of which may potentially be novel imprinted DMRs.

Next, we analysed the relationship between DNA methylation and histone modifications, both of which contribute to regulation of gene expression^{6,16}. Since human embryonic stem (ES) cells are derived from and also similar to the pluripotent ICM of blastocysts, we used the chromatin immunoprecipitation followed by sequencing (ChIP-seq) data set of the histone modifications of human ES cells compiled as part of the ENCODE Project (see online Methods)^{17–19}. Notably, we found that the H3K27me3 regions generally have low levels of DNA methylation in human ES cells and the ICM; this is also the case in gametes and cleavage-stage embryos. As a control, the regions free of H3K27me3 peaks tend to have much higher levels of DNA methylation in the ICM (Fig. 3a). This indicates that, similar to human ES cells, in the pluripotent ICM of blastocysts the genomic regions containing H3K27me3 marks probably exhibit little DNA methylation. In fact, when we analysed the DNA methylation of promoter regions of genes with H3K27me3 marks, we found that there is a strong negative correlation between the DNA methylation level of the promoter and H3K27me3 enrichment, indicating that in the pluripotent ICM, DNA methylation and H3K27me3 repress different sets of target genes (Fig. 3b and Extended Data Fig. 9a). A similar mechanism has been found during the early lineage differentiation of human ES cells^{20,21}. In contrast to H3K27me3, the H3K9me3-marked

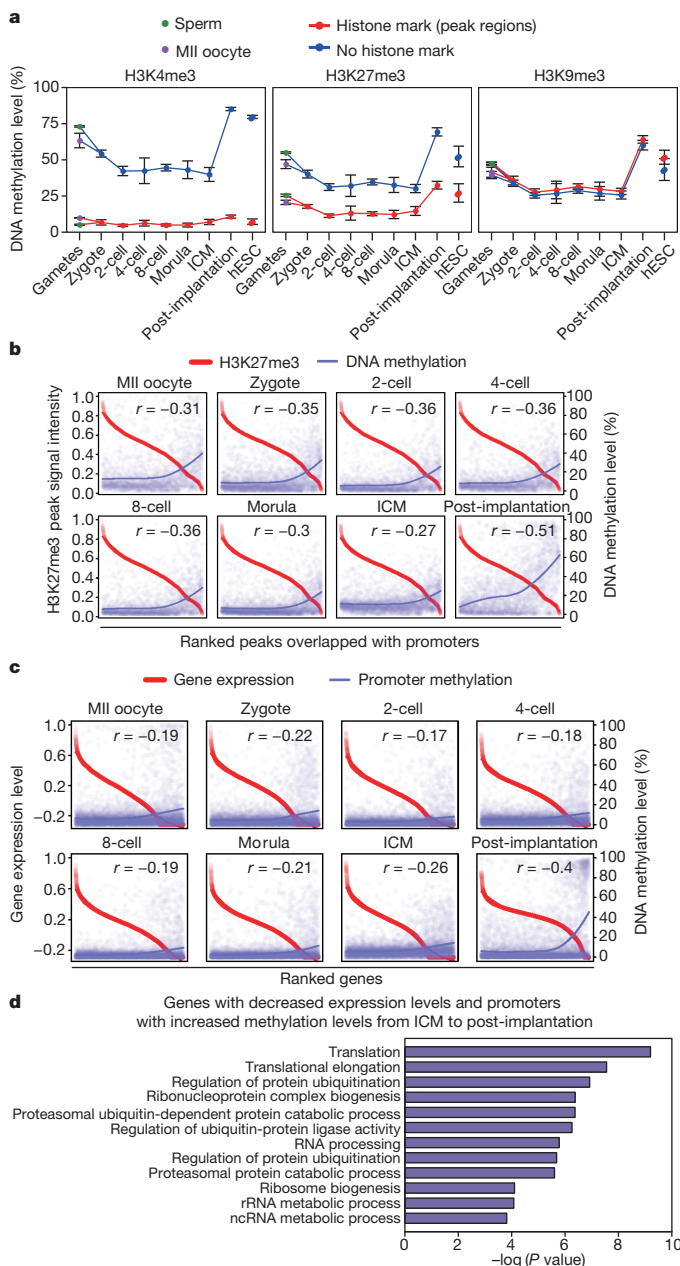


Figure 3 | Relationships between DNA methylation, histone modifications and gene expression. **a**, Average DNA methylation levels of regions with histone marks (red line) and without these histone marks (blue line) among consecutive developmental stages (details of biological replicates of each stage are listed in Supplementary Table 1) and human ES cells ($n = 2$). The left, middle and right panels represent H3K4me3, H3K27me3 and H3K9me3, respectively. The green dot indicates sperm, while the purple dot indicates MII oocytes. Data are mean \pm 95% confidence interval (± 1.96 standard error of the mean (s.e.m.)). **b**, Scatter plot of the signal intensities of H3K27me3 ChIP-seq peaks within promoter regions in human ES cells and the DNA methylation levels of the corresponding peak regions. The Pearson correlation coefficients (r) between peak signal intensities and DNA methylation levels of peak regions across every developmental stage were calculated and are included on the top right corner of each panel. The red and blue fitting curves represent peak signal intensity and DNA methylation level in corresponding regions, respectively. The horizontal axis from left to right of each box represents the H3K27me3 peaks, which overlapped with promoter regions, ranked by peak signal intensities from high to low, respectively. **c**, Scatter plot of DNA methylation levels of promoter regions and the relative expression levels of corresponding RefSeq genes. The \log_2 of the gene expression levels (reads per kilobase per million) were calculated and are presented. The Pearson correlation coefficients (r) between DNA methylation levels of promoter regions and the scaled expression levels of the corresponding genes across every developmental stage were calculated and are included in the top right corner of each panel. The red and blue fitting curves in each display represent gene expression levels and DNA methylation levels in promoter regions, respectively. The horizontal axis from left to right below each box represents the expression levels from high to low of RefSeq genes, respectively. **d**, Gene ontology analysis of the genes inactivated during implantation (fold change (ICM divided by post-implantation) ≥ 2), ($P \leq 0.05$, P value is calculated based on the multiple Student's t -test), while the methylation levels of their promoters increased. The x axis represents the negative log of the P values of the enrichment of the corresponding gene ontology terms.

regions in human ES cells exhibit similar DNA methylation levels compared with the control regions in the ICM, indicating that some overlap exists between DNA methylation and H3K9me3 in the genomes of the ICM of blastocysts, probably at constitutive heterochromatin regions (Fig. 3a and Extended Data Fig. 9b).

We also examined H3K4me3, a mark of promoters²², and found that H3K4me3 peaks tend to have very low levels of DNA methylation even in the oocytes and sperm (10% and 5%, respectively) (Fig. 3a). That is, the genes that are active, or at least permissive, in pluripotent human ES cells, and probably also in pluripotent ICM cells, are retained with minimal levels of DNA methylation in the gametes, and DNA demethylation is in general not required for these regions during early embryonic development (Extended Data Fig. 9c).

Next we investigated how DNA methylation regulates gene expression, using the single-cell transcriptome data we recently published²³. As expected, the DNA methylation levels at promoter regions negatively correlate with the expression levels of corresponding genes during early embryonic development^{16,20}. Moreover, the strength of this correlation increases gradually from the zygotic stage to post-implantation, especially after zygotic genome activation at the 8-cell stage²⁴ (Fig. 3c and Extended Data Fig. 9d). This indicates that although the genome-wide demethylation and re-methylation occur during early embryonic development, the DNA methylation at promoter regions still represses the expression of corresponding genes. It also indicates that after the zygotic genome activation at the 8-cell stage, the methylome and transcriptome of the embryos have direct corresponding relations and the repression of gene expression by DNA methylation at promoter regions is more prominent. Furthermore, during the genome-wide re-methylation following implantation, 258 genes with increased promoter methylation showed decreased RNA expression from ICM to post-implantation embryos. Gene ontology analysis showed that this set of genes was clearly enriched for the terms associated with translation and ubiquitination-related protein degradation, which indicates that the translation and protein degradation processes in proteasomes undergo fundamental changes during implantation, potentially due to DNA re-methylation (Fig. 3d). Notably, the DNA methylation on the gene bodies shows positive correlation with the expression levels of corresponding genes during pre-implantation development (Extended Data Fig. 9e, f).

Furthermore, we explored how DNA methylation contributes to the repression of transposable elements. We found that the expression levels of SINE/variable number of tandem repeats/Alu elements (SVAs) increased sharply from the 4-cell stage to morula stage (Extended Data Fig. 10a). This indicates that after the genome-wide demethylation of transposable elements at the 2-cell stage, the transcriptional activities of SVAs increase markedly for a short period but decrease back to the basal level before the genome-wide re-methylation, probably due to repression mechanisms other than DNA methylation (Extended Data Fig. 10b). SINEs, LINEs and long terminal repeats retained relatively abundant transcripts before the 8-cell stage, following which their transcripts gradually decreased to the basal level in post-implantation embryos (Fig. 4a, b and Extended Data Fig. 10c). Moreover, we found that SINEs and LINEs with different evolutionary ages show different demethylation patterns (Fig. 4a–d and Extended Data Fig. 10d–f). For example, both LINE-1 (L1) and LINE-2 (L2) belong to the LINE family of transposable elements, with L1 being evolutionarily younger than L2 (ref. 25). The younger L1 shows a higher methylation level than L2 in oocytes, whereas they show a comparable level of methylation in sperm. More importantly, we found that during the genome-wide demethylation process, the evolutionarily younger L1 retains higher levels of residual methylation than L2. L1 was also re-methylated to a higher methylation level after implantation (Fig. 4d). This pattern persists when further subdividing L1 into sub-groups, such as L1PA and L1ME with different evolutionary ages (Fig. 4e and Extended Data Fig. 10e, f). Similar patterns were also found for Alu elements and mammalian-wide interspersed repeats (MIR), two subfamilies of SINEs with younger and older evolutionary ages, respectively (Fig. 4a, b). This indicates that the evolutionarily young transposable elements with higher

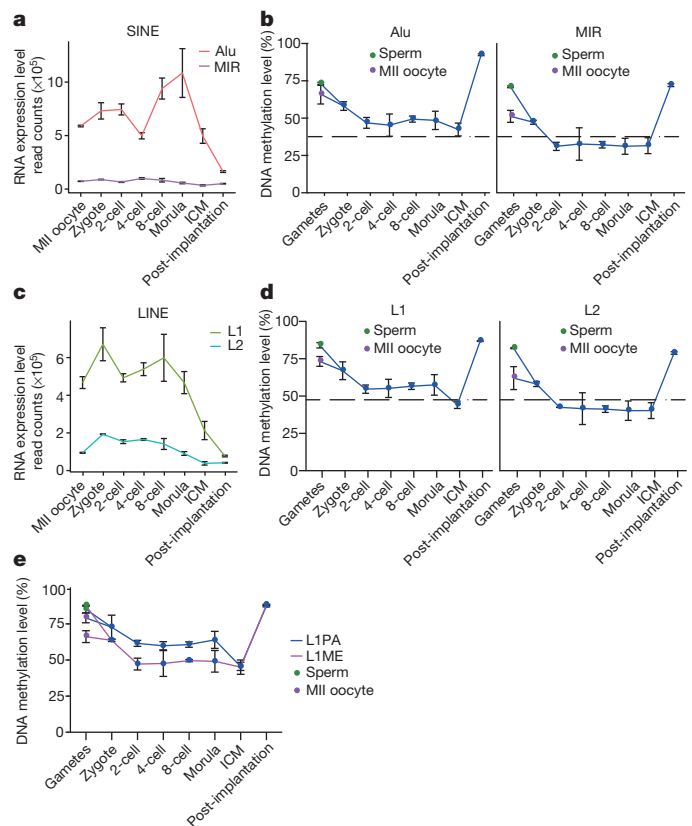


Figure 4 | Dynamics of DNA methylation and expression patterns of transposable elements. **a**, Line chart of the relative expression level (sequencing read counts, normalized by total mappable RefSeq read counts) of short interspersed nuclear elements (SINEs). **b**, Average DNA methylation levels of Alu and MIR across early embryonic stages. **c**, Line chart of the relative expression level (sequencing read counts, normalized by total mappable RefSeq read counts) of long interspersed nuclear elements (LINEs). **d**, Average DNA methylation levels of L1 and L2 across early embryonic stages. **e**, Average DNA methylation levels of two subfamilies of L1 repeat elements across early embryonic stages, with L1PA evolutionarily younger than L1ME. The green dot indicates sperm, while the purple dot indicates MII oocytes. All data in panels **a–e** are mean \pm 95% confidence interval (\pm 1.96 s.e.m.). Biological replicates in panels **a** and **c**: MII oocyte ($n = 3$), zygote ($n = 3$), 2-cell ($n = 6$), 4-cell ($n = 12$), 8-cell ($n = 20$), morula ($n = 14$), ICM ($n = 10$), post-implantation ($n = 3$); details for biological replicates of each stage in panels **b**, **d**, **e** are listed in Supplementary Table 1.

transcriptional activity tend not to be demethylated to the same extent as older elements and also retain higher levels of remnant methylation, probably to maintain stronger repression of their transcription and activity by DNA methylation (Extended Data Fig. 10).

Our work provides a comprehensive atlas at the genome-wide scale of the DNA methylation landscape in human mature gametes²⁶ and early embryos, which offers new insights that are distinct from those derived from previous findings in mice⁹.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 10 November 2013; accepted 30 May 2014.

Published online 23 July 2014.

1. Bird, A. DNA methylation patterns and epigenetic memory. *Genes Dev.* **16**, 6–21 (2002).
2. Feng, S., Jacobsen, S. E. & Reik, W. Epigenetic reprogramming in plant and animal development. *Science* **330**, 622–627 (2010).
3. Fulka, H., Mrazek, M., Tepla, O. & Fulka, J. Jr. DNA methylation pattern in human zygotes and developing embryos. *Reproduction* **128**, 703–708 (2004).
4. Li, E., Beard, C. & Jaenisch, R. Role for DNA methylation in genomic imprinting. *Nature* **366**, 362–365 (1993).

5. Hackett, J. A. & Surani, M. A. DNA methylation dynamics during the mammalian life cycle. *Phil. Trans. R. Soc. B* **368**, 20110328 (2013).
6. Lister, R. *et al.* Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature* **462**, 315–322 (2009).
7. Wu, H. & Zhang, Y. Early embryos reprogram DNA methylation in two steps. *Cell Stem Cell* **10**, 487–489 (2012).
8. Smallwood, S. A. *et al.* Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nature Genet.* **43**, 811–814 (2011).
9. Smith, Z. D. *et al.* A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**, 339–344 (2012).
10. Shirane, K. *et al.* Mouse oocyte methylomes at base resolution reveal genome-wide accumulation of non-CpG methylation and role of DNA methyltransferases. *PLoS Genet.* **9**, e1003439 (2013).
11. Seisenberger, S. *et al.* The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells. *Mol. Cell* **48**, 849–862 (2012).
12. Jaenisch, R. & Bird, A. Epigenetic regulation of gene expression: how the genome integrates intrinsic and environmental signals. *Nature Genet.* **33** (suppl.), 245–254 (2003).
13. Sasaki, H. & Matsui, Y. Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nature Rev. Genet.* **9**, 129–140 (2008).
14. Tomizawa, S. *et al.* Dynamic stage-specific changes in imprinted differentially methylated regions during early mammalian development and prevalence of non-CpG methylation in oocytes. *Development* **138**, 811–820 (2011).
15. Guo, H. *et al.* Single-cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Res.* **23**, 2126–2135 (2013).
16. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nature Rev. Genet.* **13**, 484–492 (2012).
17. Zhu, J. *et al.* Genome-wide chromatin state transitions associated with developmental and environmental cues. *Cell* **152**, 642–654 (2013).
18. The ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
19. O'Leary, T. *et al.* Tracking the progression of the human inner cell mass during embryonic stem cell derivation. *Nature Biotechnol.* **30**, 278–282 (2012).
20. Xie, W. *et al.* Epigenomic analysis of multilineage differentiation of human embryonic stem cells. *Cell* **153**, 1134–1148 (2013).
21. Gifford, C. A. *et al.* Transcriptional and epigenetic dynamics during specification of human embryonic stem cells. *Cell* **153**, 1149–1163 (2013).
22. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* **448**, 553–560 (2007).
23. Yan, L. *et al.* Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nature Struct. Mol. Biol.* **20**, 1131–1139 (2013).
24. Niakan, K. K., Han, J., Pedersen, R. A., Simon, C. & Pera, R. A. Human pre-implantation embryo development. *Development* **139**, 829–841 (2012).
25. Cordaux, R. & Batzer, M. A. The impact of retrotransposons on human genome evolution. *Nature Rev. Genet.* **10**, 691–703 (2009).
26. Molaro, A. *et al.* Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell* **146**, 1029–1041 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements F.T. and J.Q. were supported by grants from the National Basic Research Program of China (2012CB966704, 2011CB944504 and 2011CB966303) and from the National Natural Science of China (31322037, 31230047, 31271543 and 81170538). J.Q. and F.T. were supported by a grant from the Beijing Municipal Science and Technology Commission (Z131100005213006). L.Y. was supported by grants from the National Natural Science Foundation of China (81000275) and National Basic Research Program of China (2011CB944503). R.L. was supported by a grant from National Key Technologies Research and Development Program (2012BAI32B01).

Author Contributions F.T. and J.Q. conceived the experiments and supervised the project. H.G., L.Y., R.L., Y.L. J.Y., X.R., S.L., J.L., X.J., X.S., P.L., X.W., W.W., Y.W., X.L., F.G., X.W., X.F., J.Y. and L.W. performed the experiments. P.Z. and B.H. did all of the bioinformatic analyses. S.X.X., J.Q. and F.T. wrote the manuscript with help from all of the authors.

Author Information All sequencing data were deposited at the NCBI Gene Expression Omnibus (GEO) under accession number GSE49828. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.T. (tangfuchou@pku.edu.cn) or J.Q. (jie.qiao@263.net).

METHODS

Informed consents and bioethical approval. This study was approved by the Reproductive Study Ethics Committee of Peking University Third Hospital (Research license 2012SZ015). All of the gametes and embryos were collected voluntarily after obtaining the written informed consents signed by donor couples. All of the gametes and embryos were obtained from the donors at the Center for Reproductive Medicine in Peking University Third Hospital using standard clinical protocols as described^{27,28}. The donors of oocytes, zygotes, 2-cell and 4-cell embryos were financially compensated for their effort, time, inconvenience and discomfort associated with the donation process.

Preparations of human oocytes and pre- and post-implantation embryos. The average age of all donors in this study was 30 years old. All embryos used in this research have good morphology with appropriate developmental speed^{23,24}. Oocytes, sperm, zygotes, 2-cell stage and 4-cell stage embryos were voluntarily donated from the healthy volunteers who have already had one or two healthy children from natural pregnancy.

Oocytes were obtained from the donors at the Center for Reproductive Medicine in Peking University Third Hospital using standard clinical protocols as described²⁷. Embryos, at the 8-cell, morula and blastocyst stages were donated from couples who had undergone *in vitro* fertilization (IVF) treatments. These donor couples, whose infertility is purely due to female tubal factors, had a healthy baby through the IVF cycle already. They then donated the surplus frozen embryos for research with the written informed consents signed by them.

The cumulus cells around the oocytes were removed by hyaluronidase (Sigma) treatment. Only the mature MII oocytes were used in our study. Embryos produced by intracytoplasmic sperm injection (ICSI) were cultured in G1.3 medium (Vitrolife, Sweden) droplets covered by mineral oil (Sigma) until day 2. Pronuclear, 2-cell and 4-cell stage embryos were picked at the appropriate time²⁴. The first and second polar bodies were collected by a micropipette with laser-assisted biopsy, at 2 and 8 h after ICSI, respectively (Extended Data Fig. 1a).

Male and female pronuclei at different stages were collected from the zygotes at several time points according to time after ICSI precisely. Before isolation, zygotes were first immersed in G-MOPS (Vitrolife), supplemented with $5 \mu\text{g ml}^{-1}$ cytochalasin B (Sigma), and then stained with $5 \mu\text{g ml}^{-1}$ Hoechst 33342 (Invitrogen) at 37°C for 10 min. Then the visible pronuclei were carefully removed by laser-assisted biopsy (HAMILTON ZILOS-tK, CRI). Single-cell reduced representation bisulphite sequencing (RRBS) libraries of male and female pronuclei were prepared following our previously published protocol step by step¹⁵. Male and female pronuclei were further discriminated based on unsupervised hierarchical clustering of genome-wide methylation patterns of pronuclei and gametes (sperm and MII oocytes), and according to the methylation levels of the covered sperm-specific DMRs (fully methylated in the sperm but unmethylated in the oocytes) and oocyte-specific DMRs (fully methylated in the oocytes but unmethylated in the sperm), respectively (Extended Data Fig. 5a, b). The results by these two analyses matched to each other consistently.

The frozen day 3 cleavage-stage embryos were thawed by taking straws from the liquid nitrogen tank, followed by several steps to remove the cryoprotectant. Thawed embryos were transferred to G2 culture medium (Vitrolife) for further culture. The 8-cell, morula and late blastocyst stage embryos were collected at 2 h, 24 h and 72 h after thawing, respectively.

The selected embryos or oocytes were transferred into an acidic solution drop (1 ml of PBS supplemented with $1 \mu\text{l}$ of 36% HCl) to remove the zonapellucida. All of the embryos without zonapellucida were then washed several times by gentle repeated pipetting to eliminate any attached cumulus cell or polar bodies. The inner cell mass (ICM) and trophectoderm (TE) of late blastocysts were cut under a stereoscope by mechanical dissection with a glass needle (Extended Data Fig. 1a).

The post-implantation samples for RRBS are obtained from the patients with multifetal pregnancy between 6 and 10 weeks of gestation with signed informed consents from donors. Transvaginal-ultrasound-guided reduction by intrathoracic KCl injection was applied to get the 10-week-old embryo samples²⁸. The 6-week-old fresh and un-fragmented embryos for clean tissue isolation were expelled from the patients after first being treated with the mifepristone (75 mg per day for 2 days) and then 600 mg misoprostol. In order to obtain relatively pure tissues from these early embryos, the expelled embryos were first washed carefully in PBS for several times to remove all of the potential maternal contaminants. The liver tissues were carefully dissected under the microscope. The isolated relatively pure tissues for the later whole genome bisulphite sequencing (WGBS) library construction were gently transferred into a new dish, and were washed for several times with PBS to remove any excessive blood and other possible contaminations. The gestational ages of the implanted embryos used in our study were assessed by the ultrasound result and the morphology and size of the embryos.

Semen collection and preparation. Human semen samples were obtained from healthy donors by masturbation after an abstinence period of 2–3 days. After seminal liquefaction, the sample was transferred to a sterile 10 ml centrifuge tube, and washed

twice with HTF medium (Lifeglobal) supplemented with 10% human serum albumin (HSA, Vitrolife). After washing, the supernatant is removed with a sterile pipette and 2 ml HTF medium is carefully added into the tube on the top of the pellet. After swimming-up, the sperm with vigorous activity separated from the leukocytes and dead sperm were collected for DNA extraction.

Immunostaining of the oocytes and early embryos. Human zygotes and 2-cell embryos were harvested according to the time points precisely after ICSI. Human MII oocytes, zygotes and 2-cell embryos were first treated with acidic solution (pH 2.5, 1 ml of PBS supplemented with $1 \mu\text{l}$ of 36% HCl) to remove the zonapellucida. After thorough washing, the samples were fixed with 4% paraformaldehyde (Sigma) for 30 min at room temperature and washed three times in PBS, supplemented with 0.1% BSA (Sigma), and then followed by membrane permeabilization with 1% Triton X-100 (Sigma) for 2 h. For 5mC and 5hmC staining, the DNA of the oocytes and embryos were denatured with 4 M HCl at room temperature for 20 min and subsequently neutralized by 100 mM Tris-HCl buffer (pH 8.0) for 20 min, with 0.1% BSA (Sigma). After washing, samples were blocked in blocking solution, which contains 3% BSA (Sigma) in PBS. After blocking at 4°C overnight, samples were incubated with anti-5-methylcytosine antibody (1:50, BMECY-0500; Eurogentec) or anti-5-hydroxymethylcytosine antibody (1:50, 39769; Active Motif) for 2 h at 37°C . After washing three times, the samples were incubated with Alexa Fluor 568 goat anti-mouse IgG (1:500, A-11004; Invitrogen) or donkey anti-rabbit IgG-FITC (1:100, sc-2012; Santa Cruz) for 2 h at 37°C .

Since the asymmetrical distribution of histone lysine methylation is maintained in zygotes after sperm penetration, H3K9me3 was selected to mark the maternal chromatin and female pronuclei. Several of the zygotes were co-stained with H3K9me3 (1:50, 07-442; Millipore) and 5mC antibody (1:50, BMECY-0500; Eurogentec).

The final 5mC, 5hmC and H3K9me3 signals were detected using confocal microscopy (Carl Zeiss 710). All images were acquired and analysed using the ZEN imaging software (Carl Zeiss).

We did an immunofluorescence experiment of 5mC and 5hmC (Extended Data Fig. 6), which can clearly discriminate between 5hmC and 5mC. As a positive control, mouse zygotes were also co-stained for 5mC and 5hmC following the same protocol and using the same batch of antibodies. From the immunostaining results of mouse zygotes, we found that the male pronuclei, especially in the late stage zygotes (18 h post coitum), were stained with strong 5hmC signal, while the female pronuclei were mainly positive for 5mC, which is consistent with previous publications^{29–32}. We then applied the same protocol to the human zygotes with minor modifications. The 5hmC and 5mC staining signals in human zygotes are more complex than in the mouse zygotes. It seems that both male and female pronuclei have 5hmC signals in some zygotes.

Purification of genomic DNAs and total RNAs. Genomic DNAs of the post-implantation embryos or tissues, or collected sperm cells, were extracted using the DNeasy Blood and Tissue kit (Qiagen) following the manufacturer's instructions. The total RNAs of these post-implantation embryos or tissues were prepared using RNeasy Mini Kit (Qiagen) coupled with on-column DNA digestion following the manufacturer's standard protocol. The qualities of total RNAs were determined by Agilent 2100 Bioanalyzer (Agilent) before the construction of the sequencing libraries.

RRBS library preparation. All of the gametes and embryos, including MII oocytes, the first and second polar bodies, zygotes, 2-cell, 4-cell, and 8-cell embryos, morula, ICM, and trophectoderm cells (TE), were washed several times in PBS for the later RRBS libraries construction. Starting from limited numbers of embryos or single pronuclei, we integrated DNA extraction, MspI digestion, end-repair, dA-tailing, adaptor-ligation, and bisulphite-conversion into a one-tube reaction, to minimize the unnecessary DNA losses¹⁵. And we spiked in 0.5% unmethylated lambda DNA (Fermentas) to the gDNA samples before MspI digestion for monitoring the bisulphite conversion rate. After bisulphite treatment, the converted DNA was purified using regular zymo spin columns (Zymo Research) with 10 ng tRNA (Roche) as a protective carrier, followed by 2 rounds of PCR enrichments. After amplification, 200–500 bp DNA fragments were size-selected and recovered after resolving on the 12% native polyacrylamide TBE gel.

RRBS libraries of human post-implantation embryos and sperm cells were constructed following the standard RRBS protocol¹³. 100-bp pair-end sequencing was performed on Illumina HiSeq2000/HiSeq2500 platform.

For RRBS from small number or bulk of cells (first polar bodies, second polar bodies, MII oocytes, sperm, zygotes, 2-cell, 4-cell, 8-cell embryos, morulae, ICM of blastocysts, TE of blastocysts, and post-implantation embryos), in total we analysed 32 samples and generated 168.7 Gb sequencing data. For the single-cell RRBS, in total we analysed 35 samples (28 male and female single pronuclei from 14 zygotes at four different time points after ICSI, 3 single MII oocytes and 4 single sperm cells) and generated 116.4 Gb sequencing data. Note that the RRBS data set of the human ES cells was downloaded from Gene Expression Omnibus (accession number: GSM822615). Based on the RRBS analyses, we found that similar to that in mice, genome-wide demethylation happens after fertilization and the DNA methylation

level reaches the lowest level at the blastocyst stage in the ICM for both genic regions and transposable elements⁹ (Fig. 1b and Extended Data Fig. 3b). However, the major wave of demethylation is completed at the 2-cell stage and then the genome-wide DNA methylation remains relatively stable until the morula stage in humans. This is compatible with the possibility that after 2-cell stage, *DNMT1* and *UHRF1*, which are responsible for maintenance of DNA methylation, may keep the DNA methylation stable during cell divisions in the cleavage stage embryos (Extended Data Fig. 10g). Since *DNMT1* and *UHRF1* are expressed at extremely high level in the 2-cell and 4-cell stage embryos, even if only a small fraction of these proteins is localized in the nucleus, it may be enough to maintain DNA methylation level, similar to the situation in mice^{34,35}.

WGBS library preparation. The WGBS libraries were constructed following the standard protocol^{6,26}. Briefly, genomic DNA with 0.5% unmethylated lambda DNA (Fermentas) spike-in was first sonicated into 200–300-bp fragments using Covaris S2 system. After being concentrated, the sheared DNA was end-repaired, dA-tailed and ligated with pre-methylated TruSeq DNA adapters (illumina). Bisulphite conversion was conducted with MethylCode Bisulphite Kit (Invitrogen), after bisulphite conversion, the converted templates were PCR amplified and quantified using Qubit ds DNA high sensitivity dye (Invitrogen), and standard-curve-based qPCR assay (Agilent). The final quality-insured libraries were sequenced on HiSeq2000 sequencer.

For WGBS (ICM of blastocysts and post-implantation embryos), in total we analysed 2 samples and generated 146.5 Gb of sequencing data.

Both RRBS and WGBS techniques are based on bisulphite conversion of unmethylated cytosine (C), 5-formylcytosine (5fC) and 5-carboxylcytosine (5caC) into uracil (U), 5-formyluracil (5fU) and 5-carboxyluracil (5caU), whereas 5mC and 5hmC are kept intact. The methylated cytosines called by RRBS or WGBS are the sum of 5mC and 5hmC, whereas the unmethylated cytosine called are the sum of unmodified cytosine (C), 5fC and 5caC. So a CpG site we called as demethylated means that the demethylation already goes beyond 5hmC stage (that is, unmodified C, 5fC, or 5caC).

Sequencing read quality control and alignment. All of the de-multiplexed sequencing reads that passed filters were first trimmed to remove the low-quality bases and adaptor sequences. All of the RNA-seq reads were first aligned to the hg19 RefSeq reference genome (downloaded from the UCSC genome browser), and gene expression profiles of each sample were calculated using the well-defined RPKM (reads per kilobase transcriptome per million reads) method to eliminate the effects of different sequencing depths and large variation in lengths among different transcripts^{23,36}. And we only used the expression profiles of known RefSeq genes for the subsequent analysis. As for the expression profiling of human retro-elements, we used BWA tools to align all of the filtered ‘clean’ data to the human genome assembly hg19 (downloaded from the UCSC genome browser) with the command “aln -o 1 -e 60 -i 15 -q 10 -t 8”, and used the number of mappable read counts normalized with the total number of mappable RefSeq reads as the expression profile of each annotated retro-element only if these mappable reads located in the corresponding element. As for the bisulphite sequencing, including WGBS and RRBS, the remaining paired-end reads were aligned to the human reference genome (hg19, downloaded from the UCSC genome browser) using Bismark tools (version 0.7.6) with default parameters³⁷. Additionally, the 48,502 bp lambda DNA genome was rebuilt as an extra reference for later calculating the bisulphite conversion rate of each sample.

SNP identification and allele assignment. We used the WGBS data of post-implantation stage to detect the ASM events. After the alignment of sequencing data to the hg19 reference genome, we used Bis-SNP tools³⁸ to call SNPs with likelihood ratio criteria between best and second best genotype no less than 20, mapping quality of mapped reads more than 30 and base quality for genotyping more than 5. As the version of Bismark alignment tool does not support gapped alignment, we used Indel realignment function of Bis-SNP program to improve the SNP calling accuracy. We further filtered SNPs followed the default pipeline in the user-guide of Bis-SNP using the “VCFpostprocess” function. SNPs that do not occur in dbSNP database and with less than five unique reads coverage were discarded. Next, we assigned sequencing reads to each allele based on heterozygous SNPs retained using a customized Perl script.

Identification of the methylation levels of CpG and non-CpG sites. When we analysed the methylation level of each CpG site we covered, the following algorithm was applied: the number of reported C (‘methylated’ reads) divided by the total number of reported C (‘methylated’ reads) and T (‘unmethylated’ reads) at the same positions of the reference genome is calculated as the DNA methylation level of CpG site. Every CpG site with read depths ≥ 1 were summed and counted as the total CpG coverage of the sample. But when we quantify the DNA methylation level of each sample, only the CpG sites with read coverage more than five times were taken into consideration, and we applied the 100-bp-tile-based DNA methylation calling algorithm⁹. First, we binned the genome into consecutive 100-bp tiles. The number of reported C, divided by the total number of reported C and T captured in the 100-bp tiles, is interpreted as the 100-bp-tile averaged DNA methylation level.

The DNA methylation level of each sample is the average of the 100-bp tiles we have covered in our RRBS assay, while the DNA methylation level of each reported stage is the arithmetic average value of all biological replicates across each stage. The CpG density of every CpG site was calculated as the total numbers of all CpG dinucleotides located within 50 bp upstream and 50 bp downstream of this CpG site, whereas the CpG density of every 100-bp tile was then calculated as the averaged CpG density of all CpG sites located in this 100-bp tile.

For non-CpG methylation, the same strategy was applied unless indicated otherwise. As the non-CpG methylation levels are much lower than the CpG methylation levels, we used the single base resolution data, instead of the sliding-window-based approach to calculate the non-CpG methylation level, and we just average the methylation levels of all the non-CpG sites in each stage, only if they are covered more than five times.

Identification of ASM. As CpG methylation events always happen symmetrically, the bisulphite-converted reads from the forward and reverse strands covering the same symmetric CG site were combined for analysis. Only the CpG dinucleotides both covered by at least three reads on each allele (no matter the reads originated from the forward or reverse strands) were selected for the further analysis. CpG sites with ASM were identified if these CpG sites show methylation levels higher than 75% in one allele, and lower than 25% in the other allele. And the allelic read-pairs (sharing the same SNP information) harbouring at least one ASM CpG site were further extracted to calculate the methylation levels of the ASM regions. ASM regions were defined for each heterozygous SNP with at least three ASM CpG sites identified using this SNP and the methylation level of this region is higher than 75% in one allele and less than 25% in the other allele.

Identification of dynamically methylated tiles and gamete specific differentially methylated regions (DMRs). After quantifying the 100-bp-tile DNA methylation levels using 100-bp-tile-based methylation calling algorithm, we systematically compared the DNA methylation levels of 100-bp tiles which were covered in both MII oocytes and sperm. We assigned these 100-bp tiles as gamete-specific DMRs only if the methylation level of these tiles is in one type of gametes greater than 75%, while in the other type of gametes less than 25%, with a significant $P \leq 0.05$ given by multiple Student's *t*-test and a Benjamini–Hochberg false discovery rate (FDR) ≤ 0.05 . Additionally, if the methylation level differences of 100-bp tiles in each pair of consecutive stages exceed 40%, with significance in an FDR-corrected Fisher's exact test ($P \leq 0.05$, Benjamini–Hochberg FDR ≤ 0.05), these tiles were classified as the changing tiles, while the remaining tiles were considered as the stable tiles. The dynamically methylated tiles with DNA methylation levels increased or decreased across pairwise consecutive stages were assigned as increasing or decreasing tiles, respectively. Other tiles with DNA methylation levels more than 75% or less than 25% in both sperm and oocytes were assigned as the hypermethylated or hypomethylated tiles in both gametes, respectively. The hypergeometric probability distribution model was then applied to analyse which genomic region enriched most significantly in these tiles accordingly (Extended Data Fig. 7c).

We systematically searched and identified the transient methylation imprinting regions in our RRBS data set, which showed different methylation levels in gametes, and kept about 50% methylation from 2-cell embryos to blastocysts. Among the gamete-specific DMRs, especially in 12,145 oocyte-specific DMRs, about 1,400 DMRs showed transiently imprinted till blastocyst stage. On the contrary, only 140 out of the 17,473 sperm-specific DMRs are transiently imprinted. Most of these transient imprints gained further methylation after implantation (Extended Data Fig. 7d, e).

Human genomic region annotations and calculation of their DNA methylation levels. Human promoters were first classified into three classes according to their CpG densities and observed to expected CpG ratios. High-density CpG promoter (HCP), intermediate-density CpG promoter (ICP), and low-density CpG promoter (LCP) annotations were annotated as previously published²⁰ and then converted to hg19 reference build with UCSC liftOver tool. All of the enhancers we used in this study were identified by making use of published data sets^{17,18}. We integrated all of the candidate enhancers by calling the peaks of H3K4me1 in human ES cells (GSM466739) and 12 other types of human tissues¹⁷ (adipose nuclei (GSM772757), adult kidney (GSM773001), adult liver (GSM537706), duodenum mucosa (GSM621403), duodenum smooth muscle (GSM772837), pancreatic islet (GSM537642), skeletal muscle (GSM621640), stomach smooth muscle (GSM621642), fetal brain (GSM706850), fetal heart (GSM706848), fetal kidney (GSM621409), and fetal lung (GSM706853)), and then filtered out the regions which overlapped with the potential promoters (–1 kb to +0.5 kb around TSS or H3K4me3 peak regions). These putative enhancers, either poised or active, were selected for the further study. Other chromatin modification information of human ES cells, such as H3K27me3 (GSM733748), H3K9me3 (GSM1003585), H3K4me3 (GSM733657), were downloaded from the ChIP-seq data set of the ENCODE database directly¹⁸. The annotated retro-elements, such as LINEs, SINEs and LTRs and their subfamilies, were downloaded from the RepeatMasker track of the UCSC genome browser. Other regions, such as CGIs, exons and introns were downloaded from UCSC tables with hg19 track. As

previously reported, intragenic regions were covered from TSS to TES, while the intergenic regions were defined as the complement of intragenic regions in the human genome. For each annotated genomic region, the DNA methylation level was calculated as the average DNA methylation levels of all CpG sites covered within the region with more than fivefold coverage. Additionally, when quantifying the methylation level of each gene promoter region, only those promoters with at least 5 CpG sites covered were retained for further analysis.

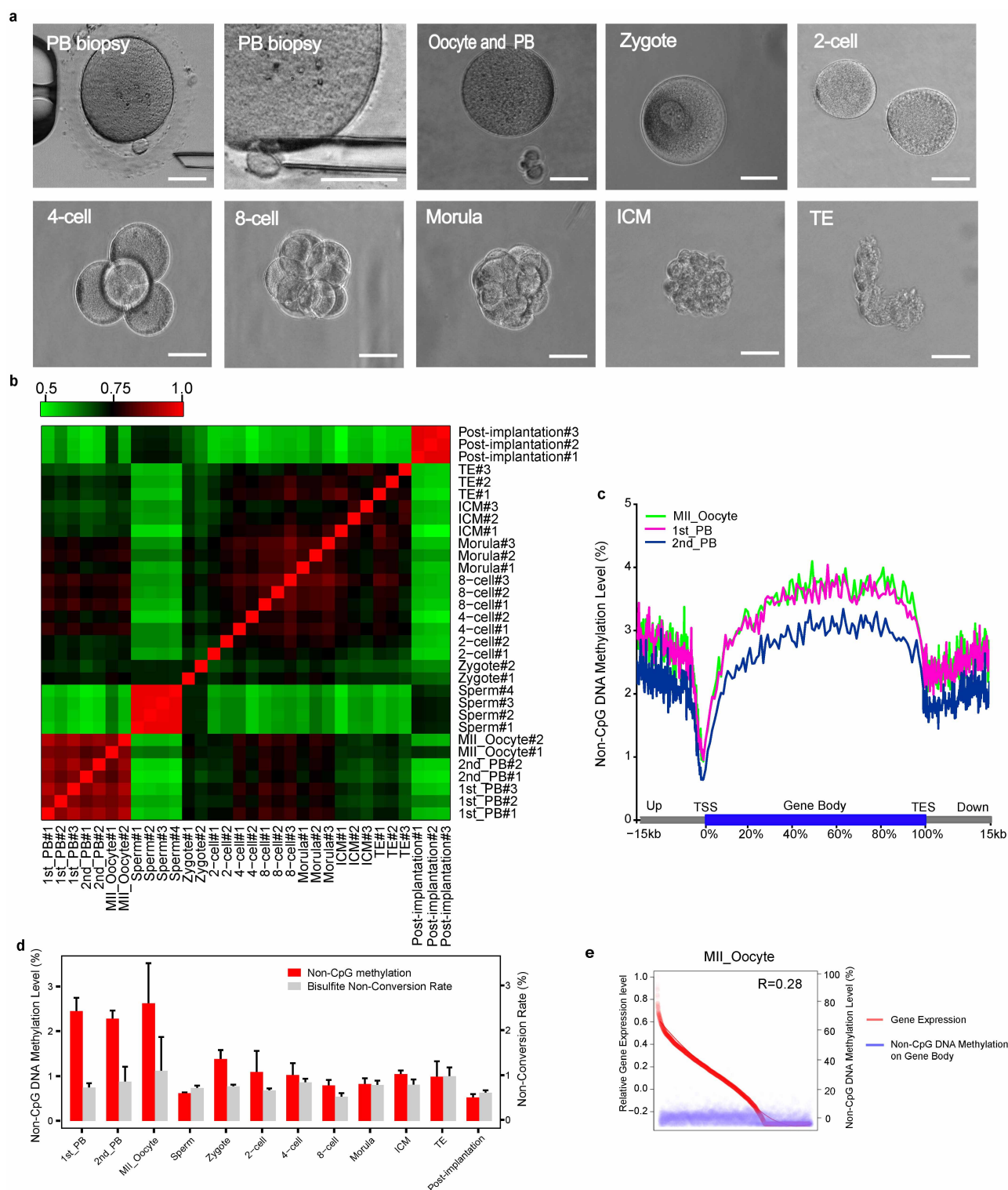
We downloaded the previously published bisulphite sequencing data of the human ES cells (Gene Expression Omnibus accession number: GSM822615), and quantified the methylation level using the same 100 bp sliding-window-based method. The methylation levels of the derived human ES cells are much higher than that in the ICM (Extended Data Fig. 9f). This is compatible with the possibility that human ES cells are more similar to post-implantation epiblast cells (similar to the mouse EpiSCs, but relatively different from mouse ESCs), although human ES cells are derived from the ICM of blastocysts. This is also compatible with our previous finding that the transcriptome of human ES cells are similar but distinct from that of the ICM (with about 1,500 genes showing differential expression between the ICM and human ES cells)²³.

For repeat elements, L1PA, L1PB, L1MA, L1MB, L1MC, L1MD and L1ME are the subfamilies within the same L1 family, with L1PA evolutionarily youngest, and L1MC, L1MD, L1ME oldest³⁹. We found that the residual DNA methylation level of L1PA is in general higher than that of L1MC, L1MD and L1ME. So it supports our conclusion that the evolutionarily younger LINEs are demethylated to a milder extent compared to older ones in the same family and have higher abundance of transcripts, implying that the early embryos tend to keep higher residual methylation at the evolutionarily younger and maybe more active transposable elements. For the subfamily of Alu, the residual methylation level of each subfamily is comparable, which is either due to the difference being too subtle to be detected under current measurement accuracy, or there is no difference between the subfamilies.

The integrated analysis of gene expression and DNA methylation. The log₂ of the gene expression levels (RPKM) of RefSeq genes were first calculated, and genes with RPKM less than 0.001 were reset to 0.001. As for the DNA methylation levels, we only displayed the methylation level of promoters of each expressed gene. The Pearson correlation coefficients (*r*) between gene expression and DNA methylation levels of promoters of corresponding genes were calculated using a customized Perl script (Fig. 3c). The gene ontology (GO) analysis was done for the following genes: (1) the expression level decreased more than twofold from ICM to post-implantation stages (*t*-test; *P* ≤ 0.05); (2) the absolute DNA methylation level of their promoters increased more than 20% (that is, level of 5mC in post-implantation – level of 5mC in ICM ≥ 20%) (Fig. 3d). The GO analysis was performed with DAVID online (<http://david.abcc.ncifcrf.gov/>)⁴⁰.

The integrated analysis of signal intensities of histone modifications and DNA methylation. We took advantage of the full ChIP-seq data set in the ENCODE database of the histone modifications in human ES cells (H3K27me3 (GSM733748), H3K9me3 (GSM1003585), H3K4me3 (GSM733657)), and made use of the signal intensity of each chromatin modification of each well-defined peak (StdSig.bigwig and broadPeak.bed files)^{17,18,20}. We simply summed the DNA methylation level of every CpG site (read depth ≥ 5) divided by the total number of the CpG sites located in every called histone modification peak as the DNA methylation level of every histone modification peak. We plotted the Pearson correlation coefficients (*r*) between peak signal intensity of each type of histone modification and DNA methylation level of each significant peak across every developmental stage using a customized Perl script (Extended Data Fig. 9a–c).

27. Li, R., Qiao, J., Wang, L., Zhen, X. & Lu, Y. Serum progesterone concentration on day of HCG administration and IVF outcome. *Reprod. Biomed. Online* **16**, 627–631 (2008).
28. Li, R. *et al.* Retain singleton or twins? Multifetal pregnancy reduction strategies in triplet pregnancies with monochorionic twins. *Eur. J. Obstet. Gynecol. Reprod. Biol.* **167**, 146–148 (2013).
29. Gu, T. P. *et al.* The role of Tet3 DNA dioxygenase in epigenetic reprogramming by oocytes. *Nature* **477**, 606–610 (2011).
30. Inoue, A. & Zhang, Y. Replication-dependent loss of 5-hydroxymethylcytosine in mouse preimplantation embryos. *Science* **334**, 194 (2011).
31. Inoue, A., Shen, L., Dai, Q., He, C. & Zhang, Y. Generation and replication-dependent dilution of 5fC and 5caC during mouse preimplantation development. *Cell Res.* **21**, 1670–1676 (2011).
32. van der Heijden, G. W. *et al.* Parental origin of chromatin in human mononuclear zygotes revealed by asymmetric histone methylation patterns, differs between IVF and ICSI. *Mol. Reprod. Dev.* **76**, 101–108 (2009).
33. Gu, H. *et al.* Preparation of reduced representation bisulfite sequencing libraries for genome-scale DNA methylation profiling. *Nature Protocols* **6**, 468–481 (2011).
34. Hirasawa, R. *et al.* Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. *Genes Dev.* **22**, 1607–1616 (2008).
35. Kurihara, Y. *et al.* Maintenance of genomic methylation patterns during preimplantation development requires the somatic form of DNA methyltransferase 1. *Dev. Biol.* **313**, 335–346 (2008).
36. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods* **6**, 377–382 (2009).
37. Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. *Bioinformatics* **27**, 1571–1572 (2011).
38. Liu, Y., Siegmund, K. D., Laird, P. W. & Berman, B. P. Bis-SNP: Combined DNA methylation and SNP calling for Bisulfite-seq data. *Genome Biol.* **13**, R61 (2012).
39. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
40. Huang da, W., Sherman, B. T. & Lempicki, R. A. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature Protocols* **4**, 44–57 (2009).

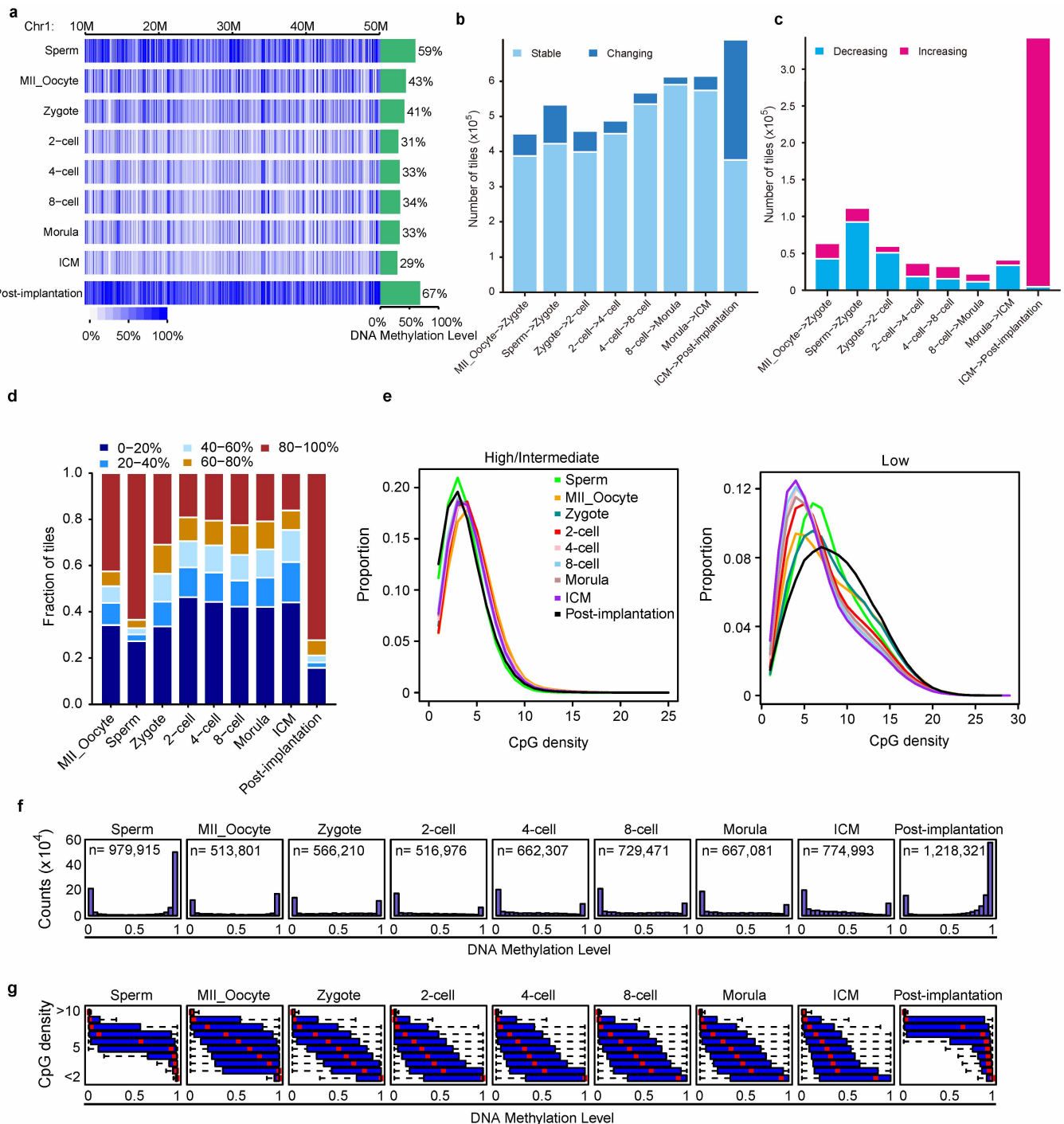


Extended Data Figure 1 | DNA methylation dynamics during human early embryonic development. **a**, Morphology of human early embryos used in this study. The microscopy images of laser assisted biopsy of polar bodies (PB), mature oocytes, and the first polar bodies, zygotes, 2-, 4-, 8-cell stage embryos, morula, inner cell mass (ICM) and trophectoderm cells (TE) from blastocyst stage embryos. Notably, the zona pellucida of all embryos, as well as mature oocytes, is removed to avoid any possible contaminants. Scale bar, 50 μ m.

b, Pearson correlation heatmap of DNA methylomes at different developmental stages of human early embryos. The numbers in the sample names indicate different biological replicates of the same developmental stages. The colour key from green to red indicates the correlation coefficient from low to high, respectively. **c**, The averaged non-CpG DNA methylation levels of MII

oocytes, the first (1st PB) and second (2nd PB) polar bodies, along the gene bodies and 15 kb upstream of the transcription start sites (TSS) and 15 kb downstream of the transcription end sites (TES) of all RefSeq genes.

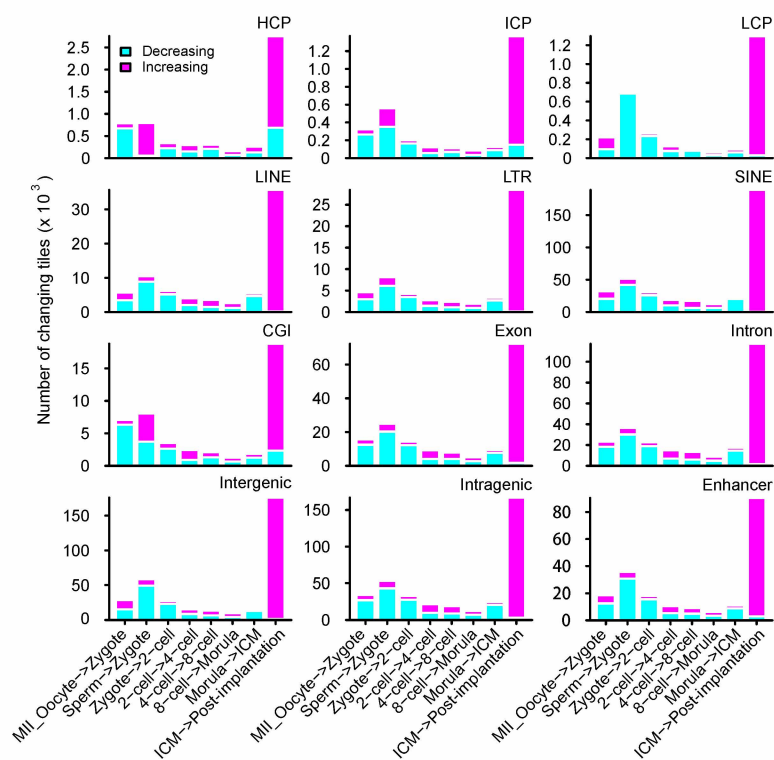
d, Non-CpG methylation levels across different human early embryonic stages. The red bars indicate the averaged DNA methylation levels of the non-CpG sites (CHG and CHH), while the grey bars indicate the bisulphite non-conversion rate of corresponding samples. All data are mean \pm 95% confidence interval (\pm 1.96 s.e.m.). Details of biological replicates of each stage are listed in Supplementary Table 1. **e**, The positive correlation of non-CpG methylation levels of the gene bodies and the expression levels of corresponding genes in MII oocytes, R indicates the Pearson correlation coefficient.



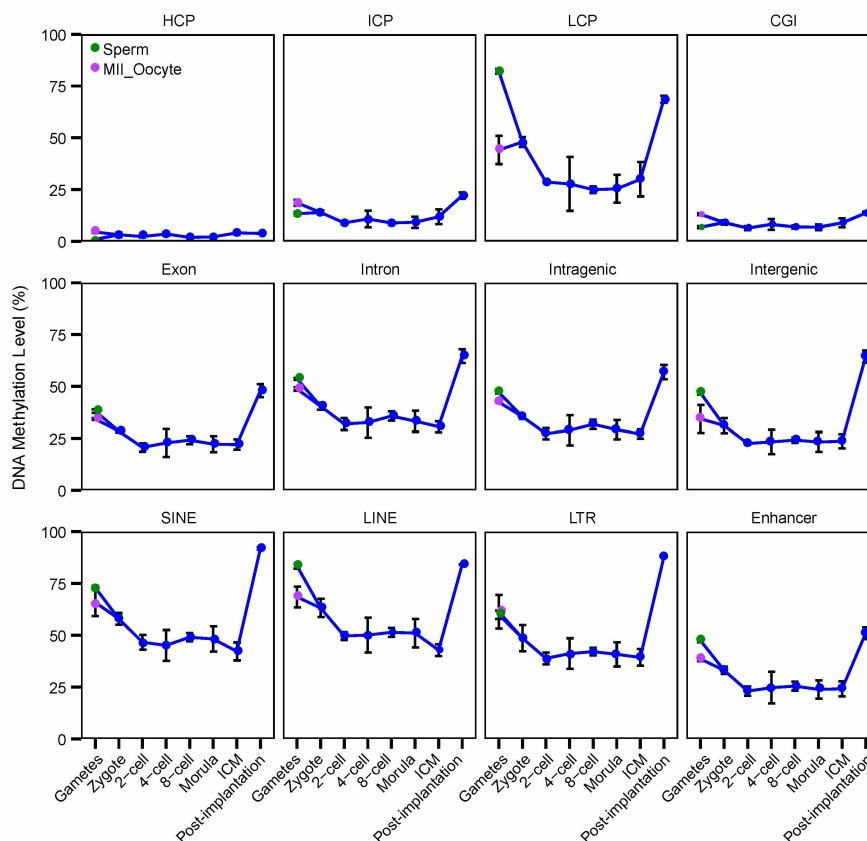
Extended Data Figure 2 | The general characteristics of the DNA methylation patterns during human early embryonic development. **a**, The heat map view of a representative section of chromosome 1 showing the dynamics of DNA methylation across different developmental stages. The green bars in the right panel indicate the average methylation levels of the corresponding regions. And the window size for the DNA methylation level calculation and presentation of these green bars is the single CpG dinucleotide covered (at least five times) in these corresponding regions. **b**, Histogram of the numbers of changing (royal blue) and stable (sky blue) tiles between consecutive stages, which shows the major transitions in DNA methylation

levels during human early embryonic development. **c**, Histogram of the numbers of DNA tiles showing increasing (magenta) and decreasing (cyan) levels of DNA methylation between consecutive stages of human early embryos. **d**, Histogram of the fractions of tiles with 0–20%, 20–40%, 40–60%, 60–80% and 80–100% methylation levels across different developmental stages. **e**, The distribution of high/intermediate (methylation level ≥ 0.2) and low (methylation level < 0.2) methylation tiles at each developmental stage against CpG density. **f**, Histogram of the counts of 100 bp tiles with different methylation levels. *n* means the total number of the 100 bp tiles for each stage. **g**, Box plots of methylation levels of each stage across local CpG densities.

a

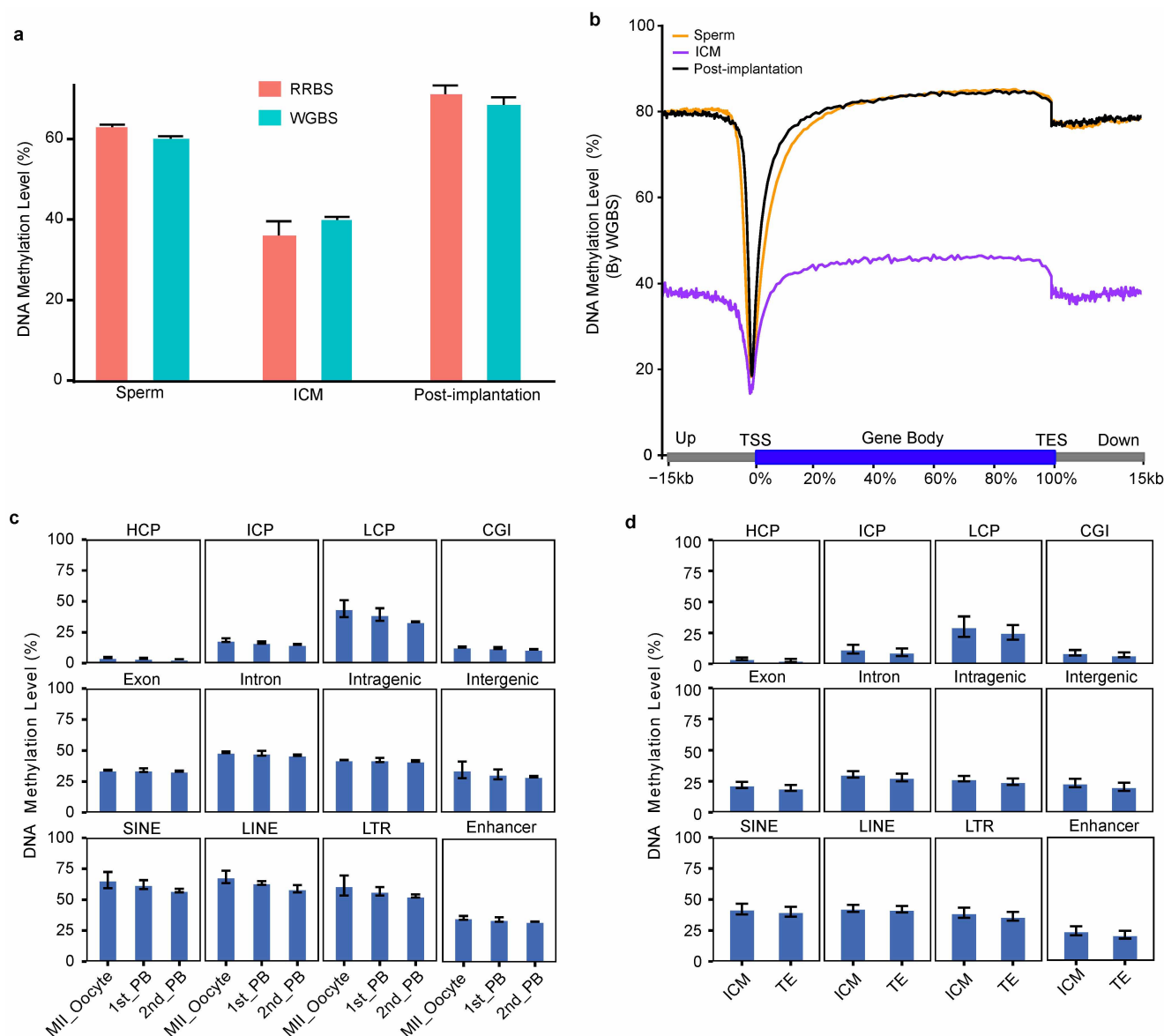


b



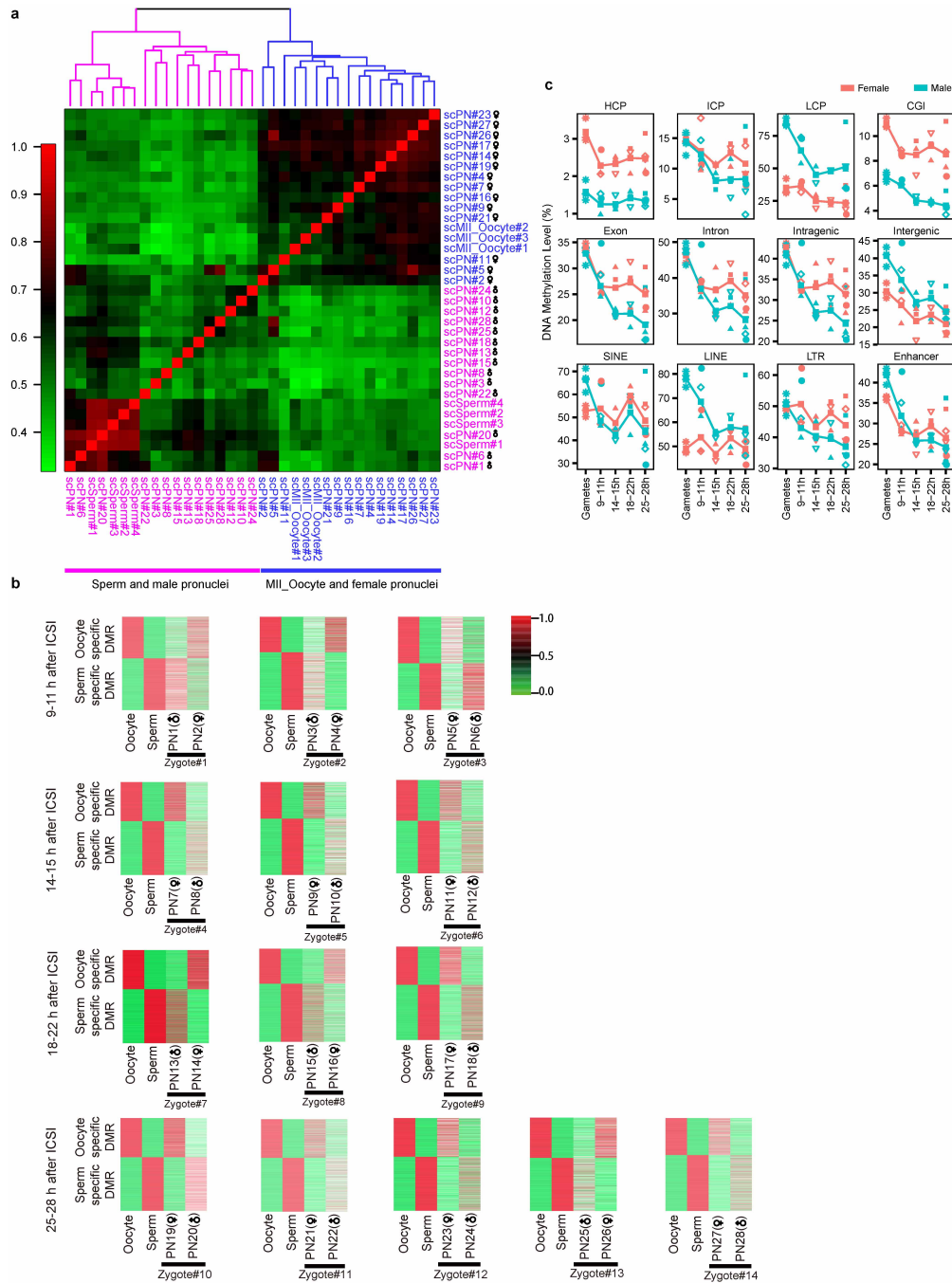
Extended Data Figure 3 | The dynamic changes of DNA methylation on a variety of annotated genomic regions. **a**, Histograms of the numbers of tiles with increasing (magenta) and decreasing (cyan) DNA methylation between each pair of consecutive stages in the annotated genomic regions during human early embryonic development. **b**, The line charts of the average methylation levels of annotated genomic regions during human early embryonic development. The green dot at gamete stage indicates the average DNA

methylation levels of the corresponding regions in sperm, while the purple dot at gamete stage indicates those in MII oocytes. HCP, high-density CpG promoter; ICP, intermediate-density CpG promoter; LCP, low-density CpG promoter; annotations as previously published²⁰. All data are mean \pm 95% confidence interval (\pm 1.96 s.e.m.). Details of biological replicates are listed in Supplementary Table 1.



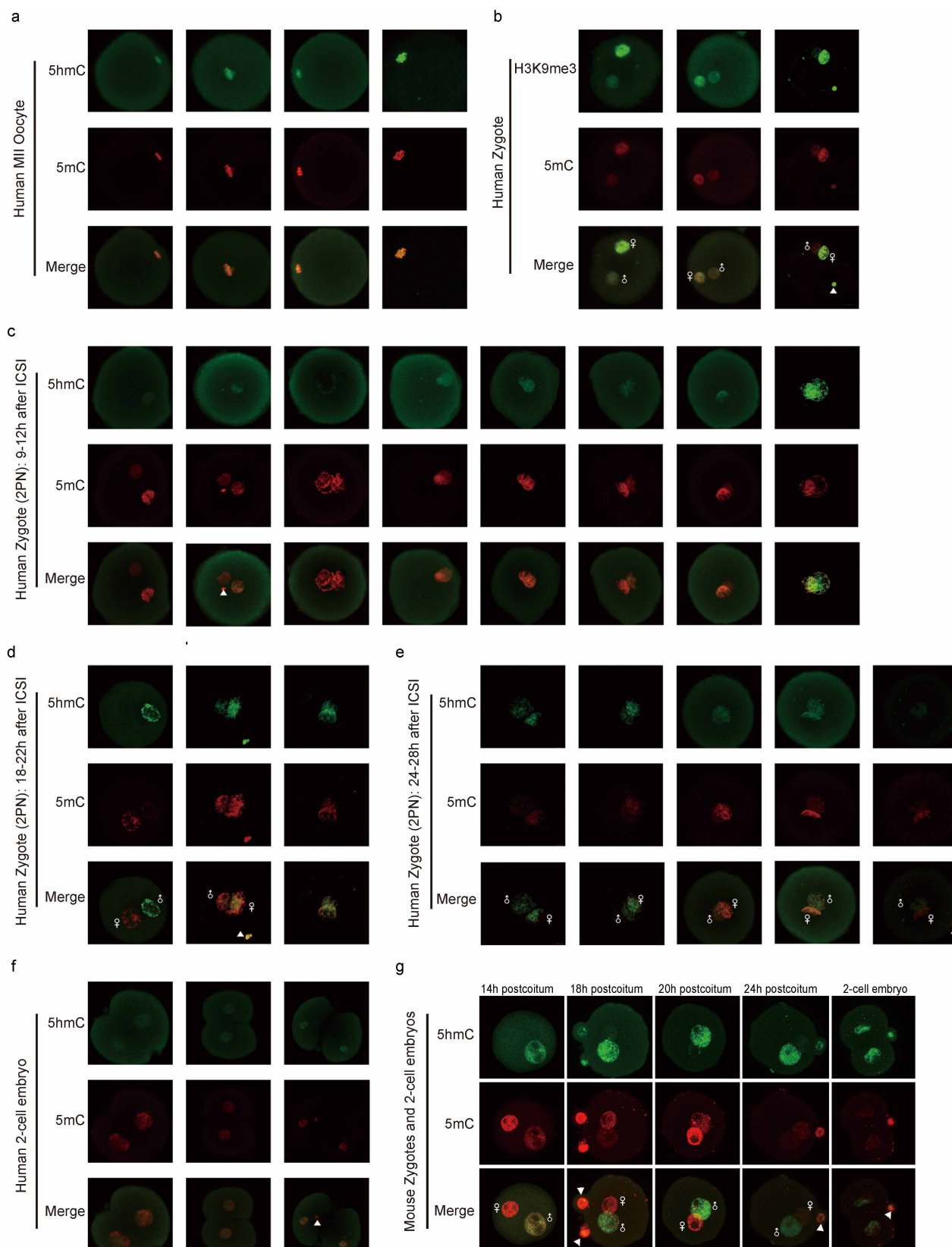
Extended Data Figure 4 | DNA methylation patterns in oocytes, polar bodies, sperm, blastocysts and post-implantation embryos. **a**, Histograms of the averaged DNA methylation levels of sperm ($n = 4$), ICM of blastocysts ($n = 3$), and post-implantation embryos (liver, $n = 3$) covered by both RRBS and WGBS data sets (WGBS data set of sperm was downloaded from Molaro, A. *et al.*²⁶). **b**, Comparison of the averaged DNA methylation levels along the gene bodies and 15 kb upstream of the transcription start sites (TSS) and 15 kb downstream of the transcription end sites (TES) of all RefSeq genes, respectively. It was analysed by WGBS for sperm (yellow line, WGBS data set

of sperm was downloaded from Molaro, A. *et al.*²⁶), ICM of blastocysts (purple line) and post-implantation embryos (liver; black line). **c**, Histograms of the average methylation levels in different genomic regions in MII oocytes ($n = 2$) as well as the first and second polar bodies (the first polar bodies, $n = 2$; the second polar bodies, $n = 2$). **d**, Histograms of the average methylation levels in different genomic regions in ICM ($n = 3$) and TE ($n = 3$) isolated from the late blastocysts. All data in panel **a**, **c** and **d** are mean \pm 95% confidence interval (± 1.96 s.e.m.).



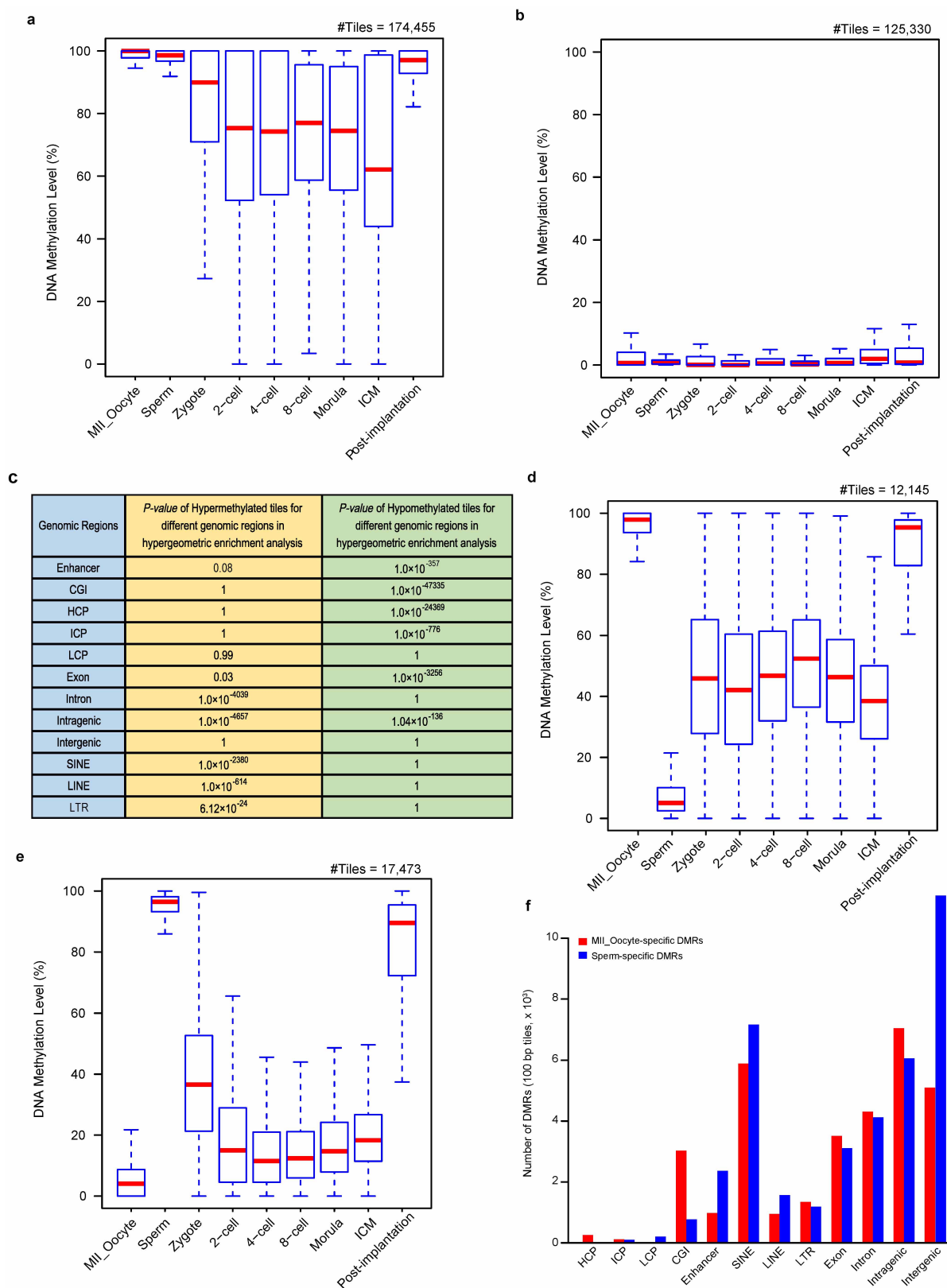
Extended Data Figure 5 | The demethylation patterns of maternal and paternal genomes on a variety of annotated genomic regions. a, Pearson correlation heatmap of DNA methylomes of individual male and female pronuclei as well as the gametes. The colour key from green to red indicates the correlation coefficient from low to high, respectively. The unsupervised clustering result shows that the single male pronuclei and sperm clustered together, while the single female pronuclei and MII oocytes clustered together. **b**, Discrimination of individual male and female pronuclei by analysing MII

oocyte-specific and sperm-specific DMRs. The sperm-specific DMRs (17,096 100 bp tiles) and MII oocyte-specific DMRs (11,850 100 bp tiles) covered by single-cell RRBS data set were used as the criterion for judging individual male and female pronuclei isolated from the same zygotes. The colour key from green to red indicates the DNA methylation levels from low to high, respectively. **c**, Demethylation dynamics of maternal and paternal genomes in human zygotes analysed by single pronucleus RRBS analysis in different annotated genomic regions.



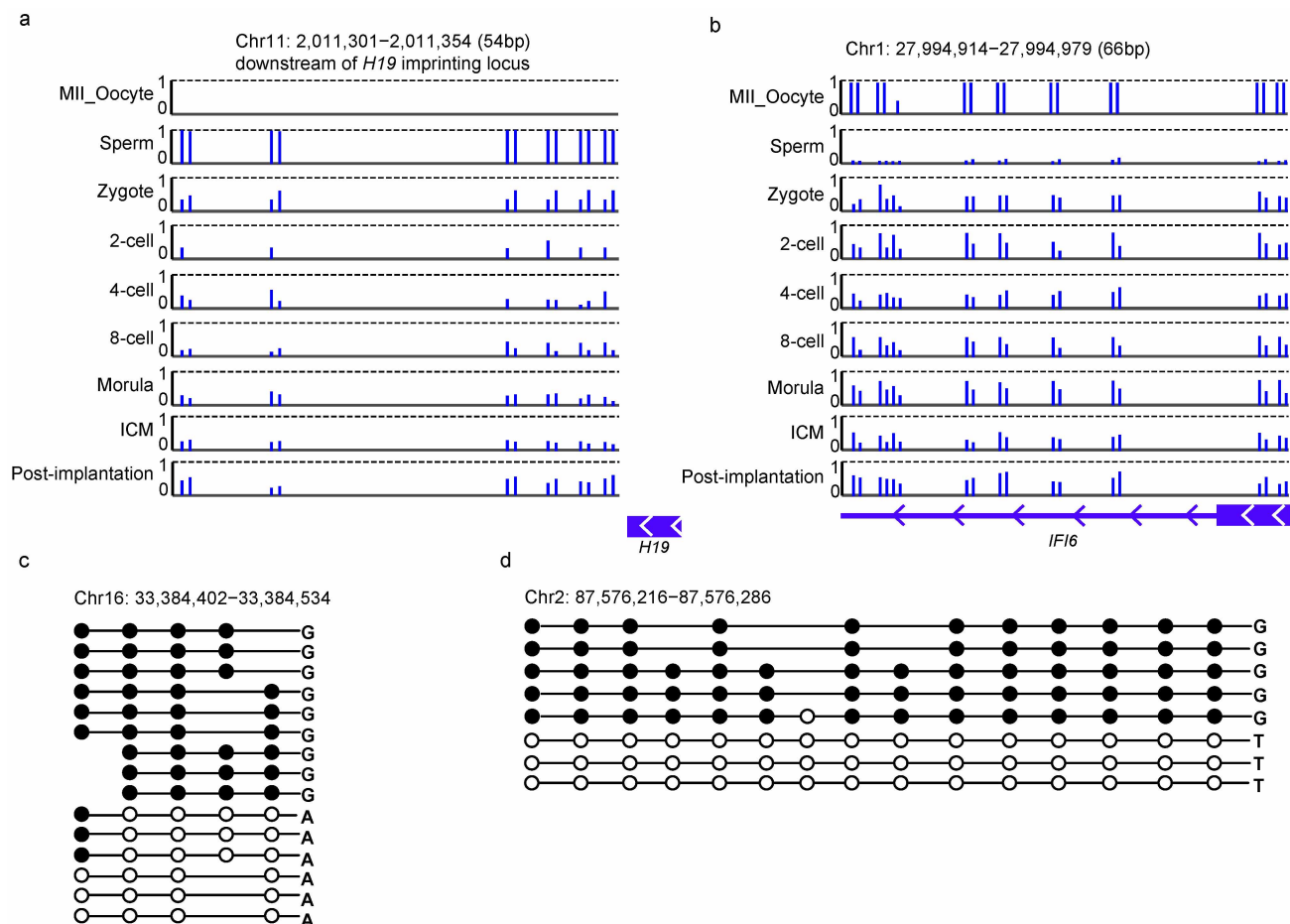
Extended Data Figure 6 | DNA demethylation patterns in pronuclear stage embryos analysed by immunostaining. a–f, The immunostaining of 5mC, 5hmC and H3K9me3 in human oocytes (a), zygotes (2PN) (b–e) and 2-cell embryos (f). The green and red signals identified from the staining indicate the 5hmC and 5mC modifications, respectively (a, c–f). Male and female symbols in the merged panels indicate the male and female pronuclei. The white triangle symbols indicate the polar bodies. b, The immunostaining of 5mC and

H3K9me3 in human zygotes. Human zygotes were co-stained with 5mC (red) and H3K9me3 (green, pronuclei with intense H3K9me3 signals were female pronuclei). g, The immunostaining of 5mC and 5hmC in mouse zygotes and 2-cell embryos as controls. The green and red signals identified from the staining indicate the 5hmC and 5mC modifications, respectively. Male and female symbols in the merged panels indicate the male and female pronucleus. The white triangle symbols indicate the polar bodies.



Extended Data Figure 7 | DNA methylation changes of DMR regions and non-DMR regions of the human gametes during pre- and post-implantation embryonic development. **a**, Box plots of DNA methylation levels for hypermethylated 100 bp tiles (average methylation levels $\geq 75\%$) in both gametes across early embryonic development stages. **b**, Box plots of DNA methylation levels for hypomethylated 100 bp tiles (average methylation levels $\leq 25\%$) in both gametes across early embryonic development stages. **c**, The hypergeometric enrichment analysis of the hypermethylated and

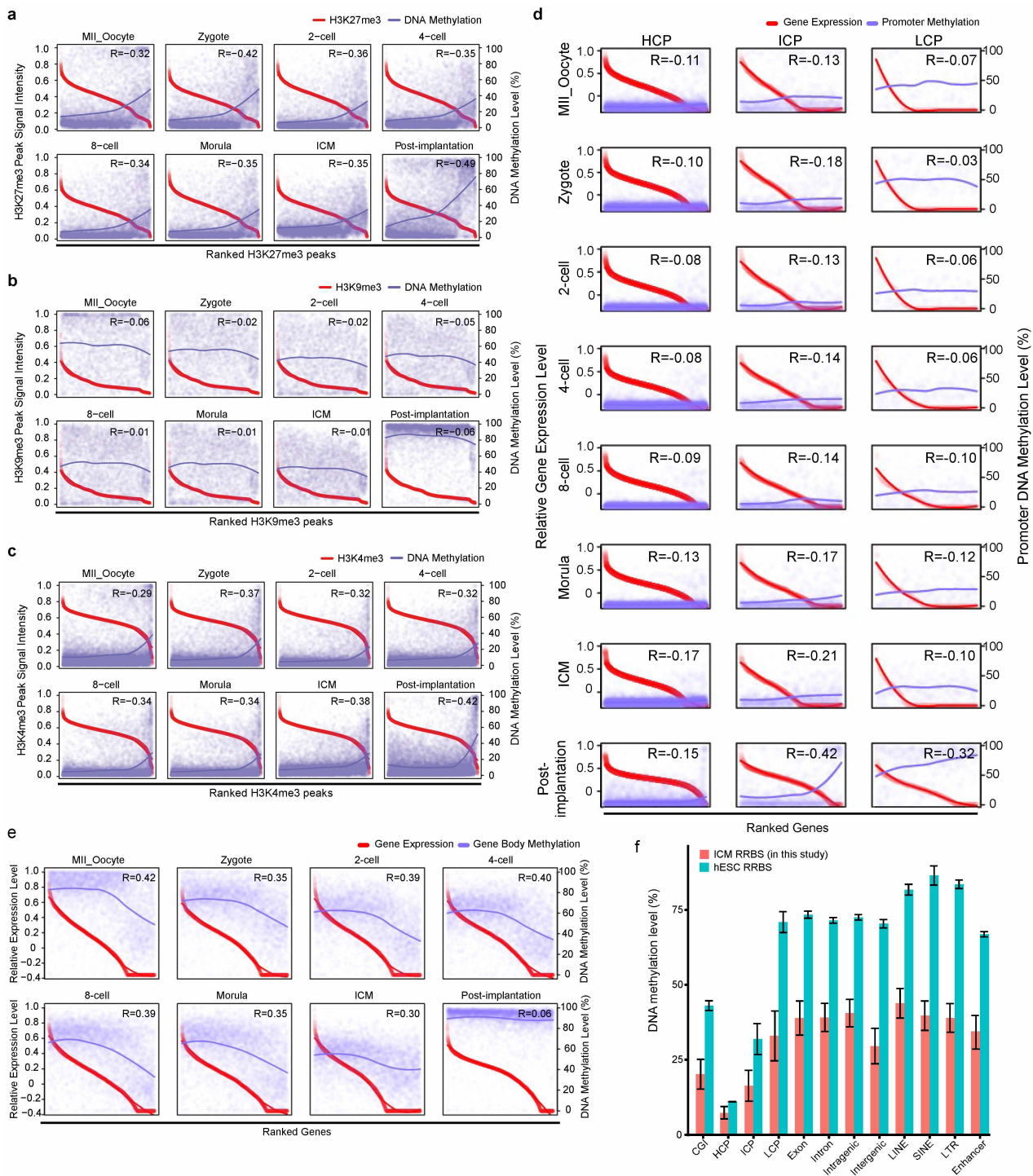
hypomethylated tiles in both gametes, exhibited the strong enrichment for different genomic regions (hypergeometric enrichment test). **d**, Box plot of methylation levels of oocyte-specific DMRs across different developmental stages. **e**, Box plot of methylation levels of sperm-specific DMRs across different developmental stages. **f**, The bar plot showing the numbers of gamete-specific DMRs located in different genomic regions, which indicates the strong enrichment for different regions in sperm-specific DMRs (blue bars) and oocyte-specific DMRs (red bars).



Extended Data Figure 8 | The DNA methylation patterns of imprinting genes and ASM regions during human early embryonic development.

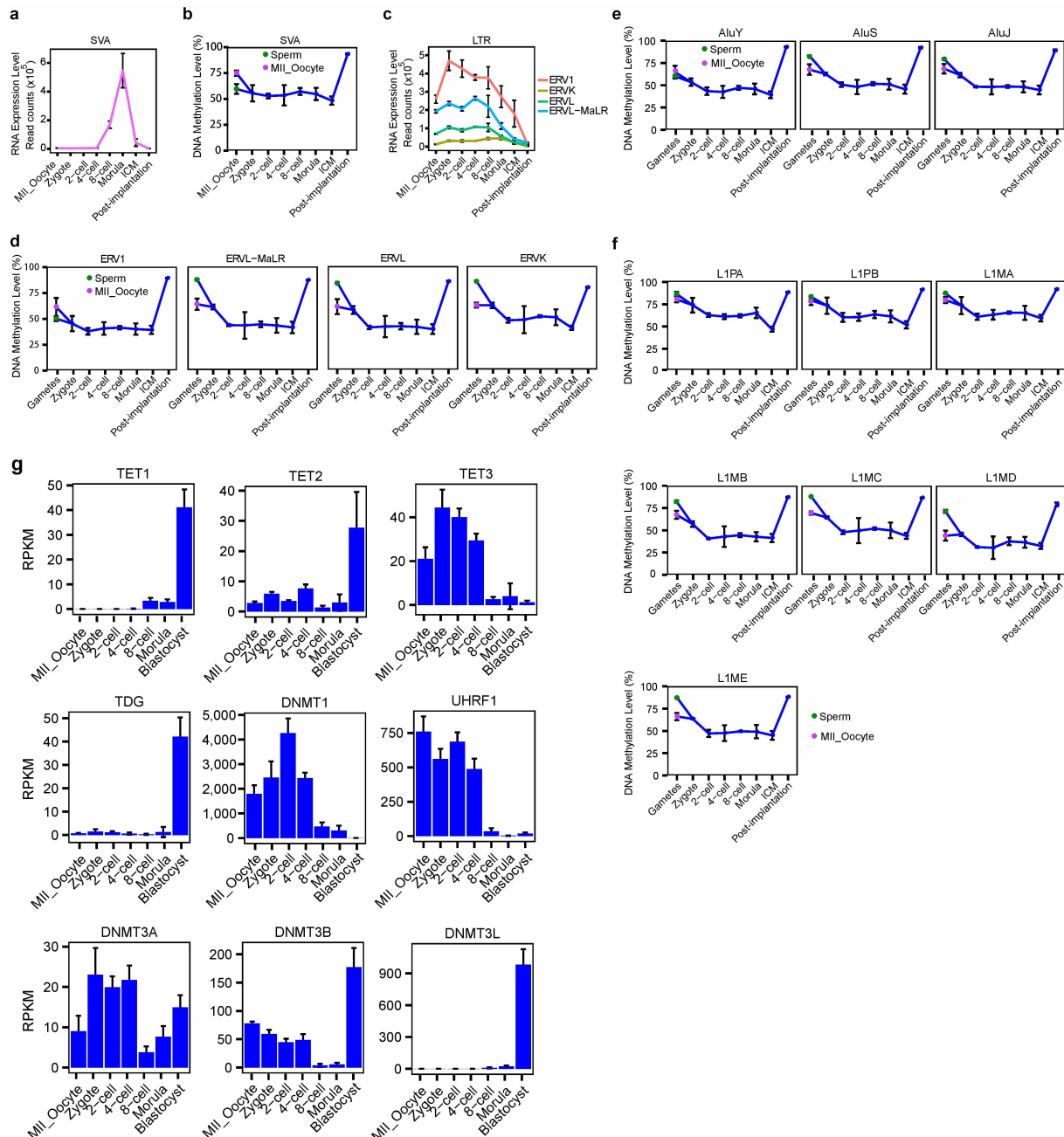
a, A representative locus of a known paternal imprinting gene, *H19*, covered in our RRBS data set. The blue bars indicate the DNA methylation levels of different CpG sites. The region was unmethylated in MII oocytes, fully methylated in sperm cells and around 50% methylated in cleavage embryos and post-implantation embryos. **b**, A representative locus of a potential novel

imprinting region within the gene body of *IFI6*, covered in our RRBS data set, which was fully methylated in MII oocytes, unmethylated in sperm, around 50% methylated in cleavage embryos and post-implantation embryos. **c**, **d**, Two allele-specific methylation loci on chromosome 16 (**c**) and chromosome 2 (**d**), tracked with heterozygous SNPs to distinguish their allele origins. The paired reads generated from the WGBS data sets with heterozygous SNPs were selected to show the DNA methylation levels of the two alleles.



Extended Data Figure 9 | The relationship of DNA methylation, histone modification and RNA expression during human early embryonic development. **a–c**, The correlation between signal intensities of three types of histone marks (H3K27me3, in panel **a**; H3K9me3, in panel **b** and H3K4me3, in panel **c**) and the DNA methylation levels of corresponding peak regions during human early embryonic development. The horizontal axis from left to right of each panel represents the peak regions of histone modifications, ranked by their signal intensities from high to low. **d**, The scatter plot of DNA methylation levels of promoter regions (HCP, ICP and LCP) and the relative expression levels of corresponding RefSeq genes. Log₂ values of the gene expression levels (RPKM) are given. The Pearson correlation coefficients (r) between DNA methylation levels of promoter regions and the scaled expression levels of the corresponding genes across different early embryonic stages were calculated and are shown in the top right corner of each panel. The red and blue fitting curves represent gene expression levels and DNA methylation levels

of corresponding promoter regions, respectively. The genes were arranged according to their expression levels. **e**, The scatter plot of DNA methylation levels of gene bodies and the relative expression levels of corresponding RefSeq genes. Log₂ values of the gene expression levels (RPKM) are given. The Pearson correlation coefficients (r) between DNA methylation levels of gene body regions (blue lines) and the scaled expression levels (red lines) of the corresponding genes across different early embryonic stages were calculated and are included in the top right corner of each panel. The red and blue fitting curves represent gene expression levels and DNA methylation levels in corresponding gene body regions, respectively. The genes were arranged according to their expression levels. **f**, Histograms of the average DNA methylation levels in different genomic regions between ICM replicates ($n = 3$) and human ES cells (GSM822615, $n = 2$) by RRBS, which showed generally higher methylation levels in human ES cells than the ICM of the blastocysts. All data are mean \pm 95% confidence interval (± 1.96 s.e.m.).



Extended Data Figure 10 | The relationship of DNA methylation and RNA expression of transposable elements during human pre- and post-implantation embryonic development.

a, The line chart of the relative expression levels (sequencing read counts, normalized by total mappable RefSeq read counts) of SVAs. Notably, the expression levels of SVAs increased dramatically from 4-cell stage to morula stage. Biological replicates in panel **a**: MII oocyte ($n = 3$), zygote ($n = 3$), 2-cell ($n = 6$), 4-cell ($n = 12$), 8-cell ($n = 20$), morula ($n = 14$), ICM ($n = 10$), post-implantation ($n = 3$). **b**, The average DNA methylation levels of SVAs. The green dot in gamete stage indicates the average DNA methylation level of the corresponding regions in sperm, while the purple dot in gamete stage indicates that in MII oocytes, respectively. Details of biological replicates of each stage are listed in Supplementary Table 1. **c**, The line chart of the relative expression levels (sequencing read counts, normalized by total mappable RefSeq read counts) of four major subfamilies (ERV1, ERVK, ERVL and ERVL-MaLR) of LTRs during early embryonic development. Biological replicates in panel **c**: MII oocyte ($n = 3$), zygote ($n = 3$), 2-cell ($n = 6$), 4-cell ($n = 12$), 8-cell ($n = 20$), morula ($n = 14$), ICM ($n = 10$), post-implantation ($n = 3$). **d**, The average DNA methylation levels of four major subfamilies of LTRs during early embryonic development. The green dot in gamete stage indicates the average DNA

methylation level of the corresponding regions in sperm, while the purple dot in gamete stage indicates that in MII oocytes. Details of biological replicates of each stage are listed in Supplementary Table 1. **e**, DNA methylation levels of the subfamilies of Alu, including AluY (the evolutionarily youngest one, the left panel), AluS (the middle panel) and AluJ (the evolutionarily oldest one, the right panel). Details of biological replicates of each stage are listed in Supplementary Table 1. **f**, DNA methylation levels of the subfamilies of L1, including L1PA (the evolutionarily youngest one in L1 family), L1PB, L1MA, L1MB, L1MC, L1MD and L1ME (the evolutionarily oldest one in L1 family). The green and red dots represented sperm and MII oocytes, respectively. Details of biological replicates of each stage are listed in Supplementary Table 1. **g**, Histograms of expression levels (RPKM) of DNA methylation-related genes across different human early embryonic stages, including DNA-demethylation-related genes *TET1*, *TET2*, *TET3* and *TDG*, as well as DNA-methylation-related genes *DNMT1*, *UHRF1*, *DNMT3A*, *DNMT3B* and *DNMT3L*. Biological replicates in panel **g**: MII oocyte ($n = 3$), zygote ($n = 3$), 2-cell ($n = 6$), 4-cell ($n = 12$), 8-cell ($n = 20$), morula ($n = 16$), blastocyst ($n = 30$). All data in panel **a-g** are mean \pm 95% confidence interval (± 1.96 s.e.m.).

DNA methylation dynamics of the human preimplantation embryo

Zachary D. Smith^{1,2,3,4,*}, Michelle M. Chan^{1,5*}, Kathryn C. Humm^{3,6,7,8,9*}, Rahul Karnik^{1,2,3}, Shila Mekhoubad^{3,4}, Aviv Regev^{1,9,10}, Kevin Eggan^{1,2,3,4,11} & Alexander Meissner^{1,2,3}

In mammals, cytosine methylation is predominantly restricted to CpG dinucleotides and stably distributed across the genome, with local, cell-type-specific regulation directed by DNA binding factors^{1–3}. This comparatively static landscape is in marked contrast with the events of fertilization, during which the paternal genome is globally reprogrammed. Paternal genome demethylation includes the majority of CpGs, although methylation remains detectable at several notable features^{4–7}. These dynamics have been extensively characterized in the mouse, with only limited observations available in other mammals, and direct measurements are required to understand the extent to which early embryonic landscapes are conserved^{8–10}. We present genome-scale DNA methylation maps of human preimplantation development and embryonic stem cell derivation, confirming a transient state of global hypomethylation that includes most CpGs, while sites of residual maintenance are primarily restricted to gene bodies. Although most features share similar dynamics to those in mouse, maternally contributed methylation is divergently targeted to species-specific sets of CpG island promoters that extend beyond known imprint control regions. Retrotransposon regulation is also highly diverse, and transitions from maternally to embryonically expressed elements. Together, our data confirm that paternal genome demethylation is a general attribute of early mammalian development that is characterized by distinct modes of epigenetic regulation.

We generated genome-scale methylation maps of human preimplantation development using reduced representation bisulphite sequencing (RRBS) to accommodate minimal DNA inputs⁷. We thawed and screened morphologically normal cleavage stage embryos and blastocysts to represent early and late preimplantation, two replicates of pooled, matched inner cell mass (ICM) and trophectoderm, motile sperm from four unrelated healthy donors, and fetal tissues (Extended Data Fig. 1). To estimate the time, extent and targets of global remethylation, we generated derivation time series of three human embryonic stem (ES) cell lines, collecting the primary outgrowth, first and fifth passage per line. On average, replicates showed high reproducibility and captured 1,753,958 CpGs of methylation data at 10× coverage (Extended Data Fig. 2).

We noted two distinguishable architectures for DNA methylation across this time series: somatic-like CpG-density-dependent bimodality in sperm, ES cells and fetal tissues, and extensive CpG-density-independent hypomethylation in preimplantation embryos (Fig. 1a). The substantial intermediate methylation in sperm reflects disparate repetitive element regulation, though non-repetitive sequences still fit the somatic paradigm¹¹ (Fig. 1b). Almost no hypermethylated CpGs persist into cleavage, with residual methylation diminishing further into the blastocyst, indicating that the embryonic landscape is rapidly established before the third embryonic division (Fig. 1b and Extended Data Fig. 2f).

Without access to human epiblasts, *in vivo* characterization of global remethylation is unavailable, but cross-species comparison between mouse epiblast and human ES cells suggest that they are a reasonably proxy for postimplantation pluripotency (Extended Data Fig. 3). Notably, within primary ES cell outgrowths, global remethylation is nearly complete, including for intermediately methylated repetitive elements apparent in sperm (Fig. 1b).

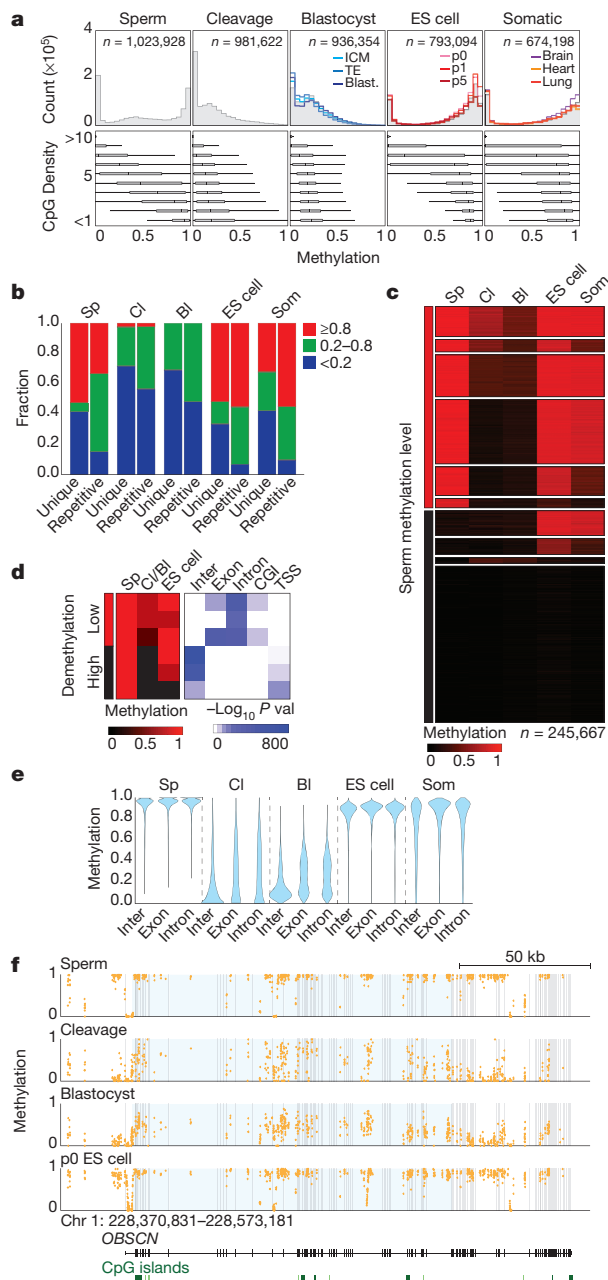
Despite predominant hypomethylation, erasure is not the *de facto* fate of all loci. Non-repetitive 100-bp genomic tiles were clustered using k-means into 10 dynamic patterns. Of sperm hypermethylated tiles, 45% retain some methylation over preimplantation, and 23% display high enough levels in cleavage embryos to be biparentally inherited and at least partially maintained (Fig. 1c). Local maintenance is significantly weighted to gene bodies: only 31% of 40,486 sperm hypermethylated intergenic tiles are ≥ 0.2 methylated in cleavage embryos, compared to 57% or 59% of exons or introns, respectively (Fig. 1d, e). Frequently, gene body methylation extends for tens to hundreds of kilobases within a single gene (Fig. 1f). Sites of retained embryonic methylation suggest residual DNA methyltransferase activity within a phase where maintenance appears otherwise impeded¹².

We incorporated recently published RNA-sequencing (RNA-seq) data to interpret the relationship between DNA methylation and expression¹³ (Supplementary Table 1). Despite global hypomethylation, the canonical negative correlation between promoter methylation and gene expression extends to preimplantation, although the overall range in promoter methylation is contracted (Extended Data Fig. 4a). Few demethylated promoters are transcribed, suggesting that promoter demethylation largely reflects the global trend (Extended Data Fig. 4b). However, demethylated promoters are more frequently induced than repressed and include *POU5F1* (also known as *OCT4*), whose embryonic induction is essential for development¹⁴ (Extended Data Fig. 4c, d). Thus, our preimplantation data support models where distinct mechanisms may regulate global versus targeted reprogramming¹⁵.

The global DNA methylation dynamics of the human embryo closely mirror those of the mouse, with sharp transitions both into and out of preimplantation^{7,16} (Fig. 2a). We investigated the behaviour of orthologous exons within the mouse preimplantation timeline⁷, predicated on their human dynamics. Surprisingly, simply sorting mouse exons according to the dynamics of their human orthologues recapitulated similar trends in sperm and over preimplantation (Fig. 2b). In mouse, demethylated exons are erased early, with moderate passive depletion over cleavage (Fig. 2c, d). Alternatively, exons that maintain methylation behave similarly in mouse and are hypermethylated in both gametes (Fig. 2c, d). Intron dynamics revealed comparable trends, but only after repetitive elements were removed from methylation estimates (Extended Data Fig. 5). Thus, both species pass through an equivalent

¹Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, USA. ²Harvard Stem Cell Institute, Cambridge, Massachusetts 02138, USA. ³Department of Stem Cell and Regenerative Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ⁴Department of Molecular and Cellular Biology, Harvard University, Cambridge, Massachusetts 02138, USA. ⁵Computational and Systems Biology Program, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. ⁶Division of Reproductive Endocrinology & Infertility, Department of Obstetrics & Gynecology, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215, USA. ⁷Obstetrics, Gynecology, and Reproductive Biology, Harvard Medical School, Boston, Massachusetts 02215, USA. ⁸Boston IVF, Waltham, Massachusetts 02451, USA. ⁹Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. ¹⁰Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. ¹¹Howard Hughes Medical Institute, Harvard University, Cambridge, Massachusetts 02138, USA.

*These authors contributed equally to this work.



global reprogramming with congruent kinetics, while many orthologous regions maintain methylation that decays passively.

Transient, maternally inherited monoallelic methylation has been previously observed in mouse^{5–7}. To identify candidate loci in human without access to oocytes, we searched for regions that are significantly more methylated in preimplantation embryos than in sperm, as this methylation would be likely to be of maternal origin (Extended Data Fig. 6a and Methods). Using our criteria, we identify 5,265 100-bp candidate maternal differentially methylated regions (DMRs), including most canonical ICRs. We clustered these regions by their resolution in ES cells and found that the majority are preimplantation-specific and either hyper- or hypomethylated in somatic tissue (Fig. 3a). The location and CpG density of DMR tiles depends on their resolution, with somatically hypomethylated DMRs substantially enriched for CpG island (CGI)-containing promoters, whereas somatically hypermethylated DMRs are more likely to be intragenic and distributed further downstream (Fig. 3a and Extended Data Fig. 6b, c). To confirm that these signatures represent true imprint-like maintenance, we generated RRBS libraries of two unrelated, single blastocysts and identified CpGs that could be assigned to each allele,

Figure 1 | Human preimplantation embryos are globally hypomethylated.

a, Top, DNA methylation across 100-bp tiles for human sperm, preimplantation embryos, including the ICM and trophectoderm (TE), ES cell derivation from outgrowth to fifth passage (p5), and somatic fetal tissues representing all germ layers. Grey highlights the average. n equals the average number of tiles captured for a given stage. The y axis shows the number of tiles in each methylation bin. Bottom, boxplots of methylation at different local CpG densities (y axis). Bullseyes indicate the median, boxes and lines the 25th and 75th, and 2.5th and 97.5th percentiles, respectively. **b**, Bar plots of 100-bp tiles segregated by non-repetitive (unique) or repetitive designation and binned by methylation status. BI, blastocyst; CI, cleavage; Som, somatic; Sp, sperm. 'ES cell' and 'Som' show the average of these time points. **c**, Non-repetitive 100-bp tiles are clustered via k -means into 10 dynamics. Sperm hypermethylated sequences follow three general trajectories: persistent maintenance, incomplete or complete demethylation. Other dynamics include sperm specific hypermethylation and hypomethylation shared by sperm and the early embryo followed by *de novo* methylation in ES cells. Finally, 3,586 tiles are hypomethylated in sperm and ES cells but methylated in embryos, representing transient imprint-like signatures. n equals the number of tiles shared between the timepoints and used for clustering. **d**, Dynamics for sperm hypermethylated, non-repetitive tiles as clustered in **c**. Left heatmap, per-cluster average of tiles. Right heatmap, $-\log_{10} P$ value of hypergeometric enrichment for each cluster for intergenic, exonic, intronic, CGI or TSS annotations using sperm hypermethylated regions as the background. **e**, Violin plot for sperm hypermethylated intergenic (Inter), exonic and intronic features. **f**, The *OBSCN* gene exhibits high inter- and intra-genic methylation and an unmethylated promoter in sperm and ES cells. In cleavage embryos, a 130-kb region, highlighted in blue, remains specifically methylated while the periphery is demethylated. Each dot refers to a CpG captured by RRBS. The y axis for DNA methylation represents the frequency in which captured CpGs are methylated and ranges from 0 to 1.

nearly all of which were monoallelically methylated (Fig. 3b and Extended Data Fig. 7).

Given that maternal DMRs in mouse are also enriched for CGIs, we next examined the conservation of targeted loci between species (Fig. 3c). We found that 795 and 293 CGIs behave as transient DMRs in human

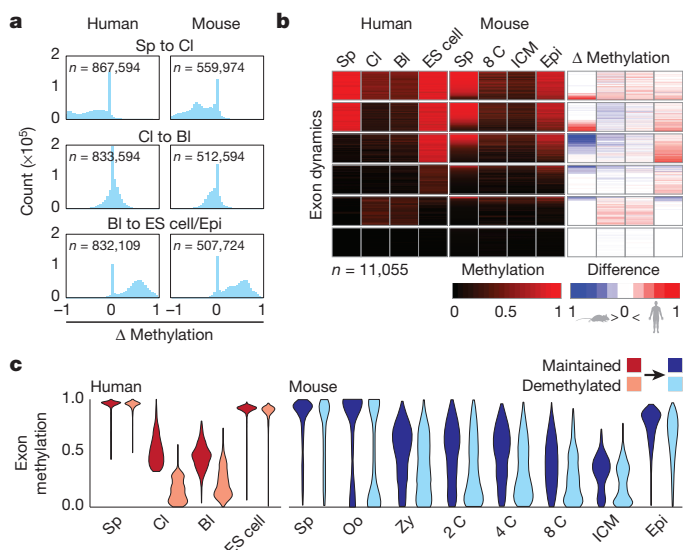


Figure 2 | Human preimplantation dynamics are globally similar to mouse.

a, Histograms of methylation changes (Δ methylation) for 100-bp tiles across human and mouse preimplantation from fertilization (Sp to CI) through preimplantation (CI to BI) to global remethylation at implantation, as measured from blastocyst to ES cell in human and ICM to embryonic day 6.5 (E6.5) epiblast in mouse (BI to ES cell/Epi). n shows the number of tiles contributing to each comparison. **b**, Exons clustered by dynamics in human with equivalent methylation values for orthologous sequences in mouse. The Δ methylation heatmap displays the difference in methylation for matched time points. n shows the number of exons available for cross species comparison. **c**, Violin plots of orthologous human sperm hypermethylated exons classified as maintained versus demethylated (top two clusters in **b**) and measured over human and mouse preimplantation.

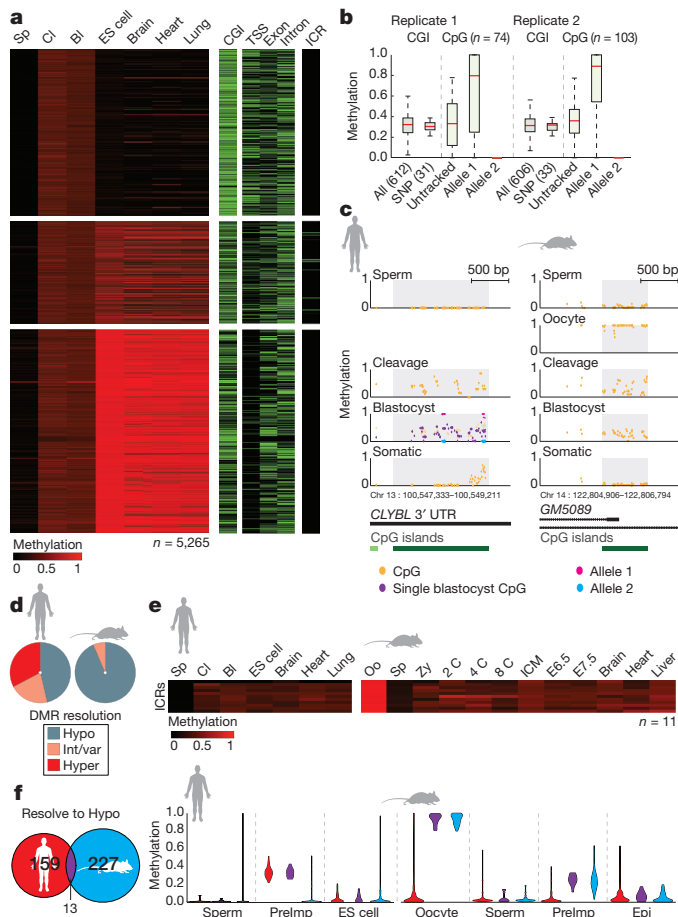


Figure 3 | Transient maternal DMRs target a divergent set of CpG island promoters. **a**, Heatmap of 5,265 100-bp tiles consistent with maternally contributed monoallelic methylation (Methods). Tiles are partitioned according to their hypomethylated (≤ 0.2), intermediate or variable (> 0.2 , < 0.8) or hypermethylated (≥ 0.8) resolution in ES cells. Feature annotations are included as separate heatmaps. **b**, Boxplots of CGI DMR methylation for two independent single blastocysts, with heterozygous SNP-linked CpGs highlighted. Within each replicate, 31 and 33 CGI DMRs contain CpGs that could be assigned to parental loci. In each case, DNA methylation is restricted to only one of the two alleles. Untracked refers to the inferred methylation status before haplotype segregation. Red line signifies the median, boxes and whiskers the 25th and 75th, and 2.5th and 97.5th percentiles. **c**, Single CpG resolution methylation values of a conserved preimplantation-specific DMR in human and mouse. Human blastocyst data includes information from the pooled sample as well as for a single blastocyst replicate (purple) with allele-tracked methylation for 10 CpGs highlighted in pink and blue. Annotated CGIs are included below. **d**, Resolution of CGIs that behave as maternal DMRs in human and mouse. **e**, Heatmap of orthologous ICRs over human and mouse preimplantation development. **f**, Orthologous hypomethylation-resolving CGI DMRs in human and mouse share only 13 equivalently regulated regions. When methylation values of mouse or human specific DMRs are tracked in the alternate species, they are constitutively hypomethylated, indicating that oogenesis targets equivalent genomic features but at species-specific sequences. PreImp refers to the average value for cleavage and blastocyst in human or 8 cell and ICM in mouse.

and mouse, respectively, with substantially more resolving to hypermethylation in human (Fig. 3d and Supplementary Table 2). Notably, human DMRs resolving to hypomethylation are more likely to be annotated as CGIs in mouse than those resolving to hypermethylation (Extended Data Fig. 6d, e). We restricted our comparison to DMRs that share CGI status and somatic hypomethylation in both species and found that maternally contributed, preimplantation-specific DMRs are strikingly divergent, with only 7.5% found in human equivalently regulated in mouse. No obvious trend distinguished shared from species-specific signatures,

though several, such as the somatic promoter of *DNMT1*, indicate conserved regulatory utility¹⁷ (Extended Data Fig. 8a). The disparity of maternal methylation targeting contrasts true ICRs, which are generally conserved¹⁸ (Fig. 3e and Extended Data Fig. 8b). Moreover, hypomethylation-resolving DMRs specific to one species are constitutively hypomethylated in the other, suggesting that these signatures are frequently and diversely targeted (Fig. 3f).

We next investigated species-specific repetitive elements, incorporating RNA-seq data to interpret DNA methylation's role in their regulation¹³. In human sperm, repetitive elements are frequently incompletely methylated or hypomethylated¹¹. Long-terminal-repeat-containing elements (LTRs) are unexpectedly bimodal, with only a fraction hypermethylated and most displaying gametic escape that persists over preimplantation (Extended Data Fig. 9a). Long interspersed nuclear elements (LINEs) are generally highly methylated in sperm, demethylated in the early embryo, with some partial remethylation in human ES cells and complete hypermethylation in somatic cells (Extended Data Fig. 9b). Finally, short interspersed nuclear elements (SINEs), which represent the majority of human genome repetitive content, exhibit a uniform global behaviour (Supplementary Table 3). Although intermediately methylated in sperm, their dynamics over preimplantation were otherwise similar to intergenic sequences in general (Extended Data Fig. 9c). Alternatively, diverse LTR and LINE subfamilies are dynamic, providing several examples where DNA methylation appears to participate in specific regulatory transitions (Supplementary Tables 4 and 5).

We found that the bimodality of LTR methylation is explained by species-specific endogenous retrovirus 1 (ERV1) family elements. Alternatively, the ERVK family generally maintains high methylation levels, similar to observations in mouse and suggesting conserved, constitutive targeting⁴ (Fig. 4a, Extended Data Fig. 9d–g). After fertilization, expression sharply transitioned from a MaIR-dominated early cleavage state resembling the oocyte to one composed of ERV1 and ERVK elements in the blastocyst and ES cells (Fig. 4b). Unexpectedly, some ERV1s seem to be induced later, including after global remethylation in ES cells, indicating a transition in the specific subfamilies that are expressed. Transcripts present early in preimplantation are generally from gametically hypomethylated ERV1 elements that are downregulated before *de novo* methylation (Fig. 4c, d). For these elements, methylation and expression are negatively correlated, indicating discriminatory targeting for even extremely related sequences (Extended Data Fig. 9h). In contrast, the LTR7 subfamily is hypermethylated in sperm, rapidly demethylated, and upregulated in the blastocyst and human ES cells (Fig. 4c, d). LTRs are rarely dynamic outside of this early versus late preimplantation axis: they are either already expressed in the oocyte and silenced later or are induced following demethylation and remain expressed in ES cells. Notably, this latter dynamic includes a limited number of recently emergent, unrelated subfamilies^{19–21}.

Compared to LTRs, LINEs maintain higher methylation levels and only the primate-specific L1A phylogeny is dynamically expressed (Fig. 5a and Extended Data Fig. 9c). As the only actively transposing lineage in humans, L1A subfamilies emerged as a linear phylogeny²². We found that the human-specific L1HS and its two closest ancestors, L1PA2 and L1PA3, are demethylated early, whereas older elements maintain higher embryonic methylation (Fig. 5b). Correspondingly, nearly all embryonic transcription could be attributed to these three youngest subfamilies (Fig. 5c and Extended Data Fig. 10a). Given the homology between subfamilies, we searched for sequence composition changes that may explain the preimplantation-specific escape of younger elements. We aligned 5' untranslated regions (UTRs) of full-length L1PA7 to L1HS and compared sequences that demarcated demethylated L1PA3-descended elements from constitutively targeted ancestors. The largest discrete difference corresponds to an ~130-bp deletion found within the L1PA3 lineage itself that separates older elements from the L1HS progenitor L1PA3a²³ (Fig. 5d and Extended Data Fig. 10b, c). Intriguingly, the presence or absence of this region isolates two disparately regulated subpopulations and marks the transition to embryonic expression (Fig. 5e, f

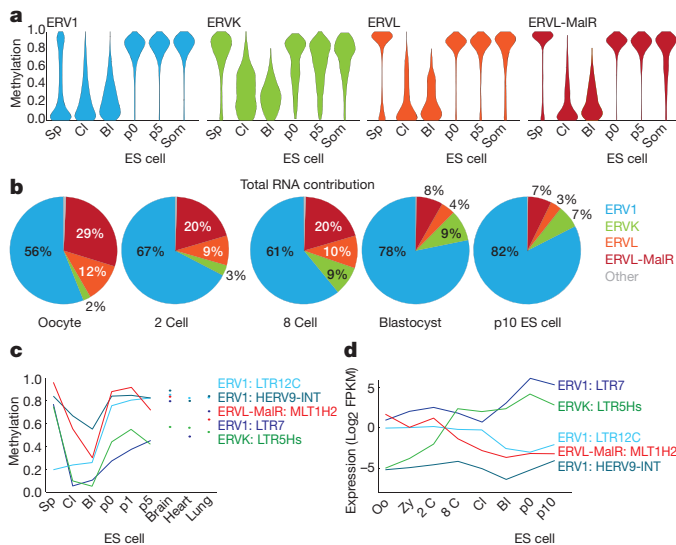


Figure 4 | LTR subfamily dynamics are divided into early and late preimplantation phases. **a**, Violin plots for the four LTR families present in human over early development and ES cell derivation. **b**, Pie charts of LTR family expression calculated as the number of fragments per million (FPM) that align to elements within the family. **c**, Mean methylation of notable ERV1, ERVK and MaIR subfamilies. Three ERV1 subfamilies are included to represent discrete dynamics: shared gamete and early embryonic hypomethylation (LTR12c), constitutive methylation (HERV9-INT) and rapid demethylation (LTR7). The ERVK subfamily LTR5Hs is also demethylated. **d**, Expression dynamics for the same subfamilies in **c**. LTR12c is expressed early and downregulated in the blastocyst. Alternatively, LTR7 is expressed throughout, but upregulated in the blastocyst and maintained in ES cells, where it accounts for the majority of ERV1 transcripts. Like LTR7, LTR5Hs is only intermediately methylated during ES cell derivation and is embryonically induced. Alternatively, the ERV1 HERV9-INT remains repressed. MLT1H2 is the prevailing MaIR transcribed in the oocyte and is lost after fertilization. Expression is the fragments per million that align to subfamily elements, divided by the kb annotated as the subfamily in the genome (FPKM).

and Extended Data Fig. 10d). This adaptation may represent a specific moment in the evolutionary progression of the LIPAs when emerging elements evaded a seemingly sequence-directed, repressive mechanism. Whether older subfamilies retain expression or transposition potential without active silencing, or if this signature reflects a vestigial, host genome adaptation that is no longer required, remains to be investigated.

We present base pair resolution maps of DNA methylation as it is dynamically reconfigured during human early development. These data identify a set of transient, maternally contributed methylation at CGI promoters, the resolution of which suggest independent modes of acquisition: male germline specific protection against methylation or *de novo* targeting in the oocyte of otherwise canonically unmethylated CGIs²⁴. Both are common and show poor conservation compared to classic ICRs, indicating that short-lived, parent-specific signatures are less evolutionarily constrained than those persisting after implantation. We find that repetitive element regulation is notably diverse in human, more so than in mouse, with gametically hypomethylated LTR subfamilies present in the oocyte and early embryo and others sharply demethylated and induced embryonically. For LINEs, the stepwise phylogeny within the primate-specific L1PA lineage pinpoints a specific, adaptive transition. As L1HS elements remain transpositionally active, including somatically in numerous cancers, the targeting of epigenetic silencing machinery during preimplantation may be relevant in identifying the root cause of their aberrant induction later²⁵. Understanding the regulatory principles inherent to the early embryo will improve continuing efforts to evaluate complex traits with unclear modes of epigenetic inheritance. Future work to characterize the mechanisms that impose these diversely targeted embryonic methylation patterns will illuminate their contribution to normal human development and disease.

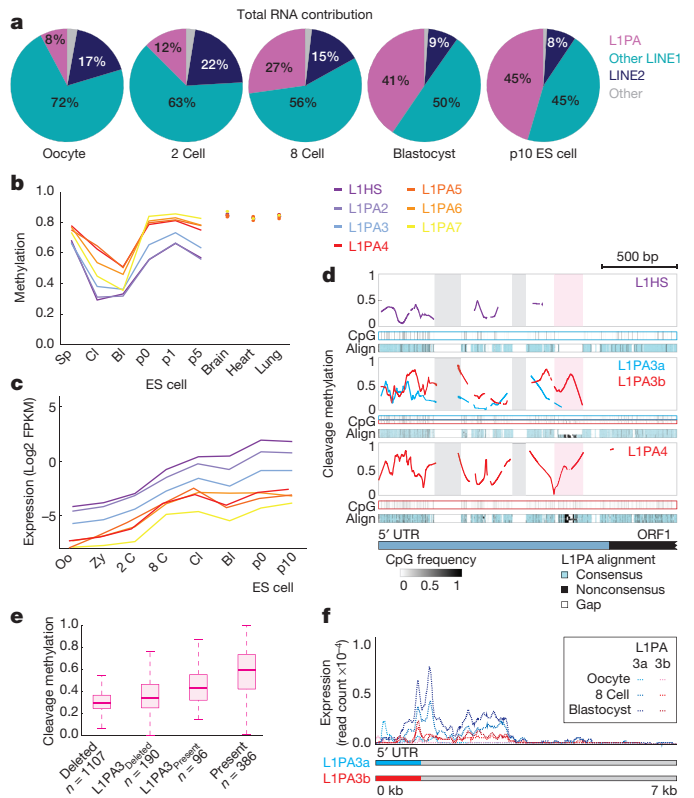


Figure 5 | Emergent L1PA subfamilies escape DNA methylation-associated repression during preimplantation. **a**, Pie charts of the LINE expression divided into the L1PA subfamily, other LINE1 and LINE2 subfamilies. Total expression is calculated as the number of fragments per million (FPM) that align to family elements. **b**, Mean methylation values for the most recent L1PA subfamilies. In cleavage embryos, L1HS through L1PA3 are demethylated and maintain these levels through the blastocyst. **c**, Expression dynamics for the same subfamilies in **c** over preimplantation and in ES cells. The three youngest L1PA subfamilies are induced by the 8 cell stage. Expression is the number of fragments per million that align to subfamily elements, divided the kb annotated as the subfamily in the genome (FPKM). The colours in **c** are the same as in **b**. **d**, Composite plot of cleavage stage methylation values across aligned 5' UTRs in L1PA subfamilies. The composite for L1PA3 is split by the presence (red) or absence (blue) of a ~130-bp sequence that distinguishes L1PA3b from L1PA3a and demarcates methylation values between older and newer subfamilies (highlighted in pink). Multiple sequence alignment for each subfamily to the assembled consensus is below each composite, with blue corresponding to conservation, black to divergence, and white to gaps or deletions. The x axis represents position along the L1HS 5' UTR and a portion of ORF1. CpG Frequency describes per CpG conservation within single elements to the consensus. Two older sequences specific to L1PA7 are highlighted in grey. **e**, Boxplot of L1PA methylation in cleavage embryos, sorted by the presence of the ~130 bp sequence for all elements and L1PA3 specifically. Preimplantation methylation is higher for elements that contain this insert. Bold line signifies the median, boxes and whiskers the 25th and 75th, and 2.5th and 97.5th percentiles. **f**, Expression composite of full-length insert deleted L1PA3a and insert containing L1PA3b subfamilies in oocyte, 8 cell and blastocyst stage embryos. Transcriptional induction is not apparent until after fertilization and is specific to L1PA3a. Read count is the read coverage normalized by total reads (Methods).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 16 September 2013; accepted 16 June 2014.

Published online 23 July 2014.

1. Suzuki, M. M. & Bird, A. DNA methylation landscapes: provocative insights from epigenomics. *Nature Rev. Genet.* **9**, 465–476 (2008).
2. Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* **480**, 490–495 (2011).

3. Ziller, M. J. *et al.* Charting a dynamic DNA methylation landscape of the human genome. *Nature* **500**, 477–481 (2013).
4. Lane, N. *et al.* Resistance of IAPs to methylation reprogramming may provide a mechanism for epigenetic inheritance in the mouse. *Genesis* **35**, 88–93 (2003).
5. Smallwood, S. A. *et al.* Dynamic CpG island methylation landscape in oocytes and preimplantation embryos. *Nature Genet.* **43**, 811–814 (2011).
6. Kobayashi, H. *et al.* Contribution of intragenic DNA methylation in mouse gametic DNA methylomes to establish oocyte-specific heritable marks. *PLoS Genet.* **8**, e1002440 (2012).
7. Smith, Z. D. *et al.* A unique regulatory phase of DNA methylation in the early mammalian embryo. *Nature* **484**, 339–344 (2012).
8. Fulka, H., Mrazek, M., Tepla, O. & Fulka, J. Jr. DNA methylation pattern in human zygotes and developing embryos. *Reproduction* **128**, 703–708 (2004).
9. Santos, F. *et al.* Evaluation of epigenetic marks in human embryos derived from IVF and ICSI. *Hum. Reprod.* **25**, 2387–2395 (2010).
10. Lepikhov, K. *et al.* Evidence for conserved DNA and histone H3 methylation reprogramming in mouse, bovine and rabbit zygotes. *Epigenetics Chromatin* **1**, 8 (2008).
11. Molaro, A. *et al.* Sperm methylation profiles reveal features of epigenetic inheritance and evolution in primates. *Cell* **146**, 1029–1041 (2011).
12. Hirasawa, R. *et al.* Maternal and zygotic Dnmt1 are necessary and sufficient for the maintenance of DNA methylation imprints during preimplantation development. *Genes Dev.* **22**, 1607–1616 (2008).
13. Yan, L. *et al.* Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nature Struct. Mol. Biol.* **20**, 1131–1139 (2013).
14. Wu, G. *et al.* Establishment of totipotency does not depend on Oct4A. *Nature Cell Biol.* **15**, 1089–1097 (2013).
15. Hackett, J. A. & Surani, M. A. DNA methylation dynamics during the mammalian life cycle. *Phil. Trans. R. Soc. Lond. B* **368**, 20110328 (2013).
16. Borgel, J. *et al.* Targets and dynamics of promoter DNA methylation during early mouse development. *Nature Genet.* **42**, 1093–1100 (2010).
17. Mertineit, C. *et al.* Sex-specific exons control DNA methyltransferase in mammalian germ cells. *Development* **125**, 889–897 (1998).
18. Smits, G. *et al.* Conservation of the *H19* noncoding RNA and *H19-IGF2* imprinting mechanism in therians. *Nature Genet.* **40**, 971–976 (2008).
19. Kelley, D. & Rinn, J. Transposable elements reveal a stem cell-specific class of long noncoding RNAs. *Genome Biol.* **13**, R107 (2012).
20. Fuchs, N. V. *et al.* Human endogenous retrovirus K (HML-2) RNA and protein expression is a marker for human embryonic and induced pluripotent stem cells. *Retrovirology* **10**, 115 (2013).
21. Lu, X. *et al.* The retrovirus HERVH is a long noncoding RNA required for human embryonic stem cell identity. *Nature Struct. Mol. Biol.* **21**, 423–425 (2014).
22. Khan, H., Smit, A. & Boissinot, S. Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates. *Genome Res.* **16**, 78–87 (2006).
23. Boissinot, S., Entezam, A. & Furano, A. V. Selection against deleterious LINE-1-containing loci in the human lineage. *Mol. Biol. Evol.* **18**, 926–935 (2001).
24. Kelsey, G. & Feil, R. New insights into establishment and maintenance of DNA methylation imprints in mammals. *Phil. Trans. R. Soc. Lond. B* **368**, 20110336 (2013).
25. Hancks, D. C. & Kazazian, H. H., Jr. Active human retrotransposons: variation and disease. *Curr. Opin. Genet. Dev.* **22**, 191–203 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We would like to thank all members of the Meissner, Regev and Eggan laboratories, in particular M. Ziller for critical reading of the text and K. Koszka for supervising human embryo thawing. We also thank D. Sakkas and R. Holmes of Boston IVF for clinical assessment of embryo morphology and viability, as well as S. Levine, M. Gravina and K. Thai from the MIT BioMicro Center. Finally, we thank T. S. Mikkelsen, H. Gu, and A. Gnirke from the Broad Institute for their guidance and expertise. A.R. is an investigator of the Merkin Foundation for Stem Cell Research at the Broad Institute. This work was supported by the Harvard Stem Cell Institute (K.E.), and NIH Pioneer Award (5DP10D003958), the Burroughs Wellcome Career Award at the Scientific Interface and HHMI (to A.R. and K.E.), P01GM099117 (to A.M. and K.E.) and a Center for Excellence in Genome Science from the NHGRI (1P50HG006193-01, to A.R. and A.M.). A.M. is a New York Stem Cell Foundation Robertson Investigator. This research was supported by The New York Stem Cell Foundation.

Author Contributions Z.D.S., K.E. and A.M. conceived the study and Z.D.S., M.M.C., K.C.H., A.R., K.E. and A.M. facilitated its design. Z.D.S., K.C.H. and S.M. collected samples and Z.D.S. performed methylation profiling, M.M.C. and R.K. performed all analysis with assistance from Z.D.S. Z.D.S., M.M.C. and A.M. interpreted the data and wrote the paper with the assistance of the other authors.

Author Information RRBS data is deposited at the Gene Expression Omnibus under accession number GSE51239. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.M. (alexander_meissner@harvard.edu) or K.E. (keggan@scr.broad.harvard.edu).

METHODS

IRB approval. Harvard University institutional review board (IRB) and Embryonic Stem Cell Research Oversight (ESCRO) Committee approval was obtained for both the collection and experimental use of surplus embryos resulting from infertility treatment and donated for research.

The Harvard University and Beth Israel Deaconess Medical Center institutional review boards both determined that collection, preparation, and experiments using discarded human gametes was not human subjects research and therefore did not require a full IRB review.

Consent process. All embryos used in this study had previously been donated and stored at Harvard University. Couples donating surplus embryos for the purpose of research signed an extensive consent form at the time of their donation. These consent forms were approved by the Harvard University IRB. The authors did not have access to any identifying personal health information.

Gametes were obtained from patients who signed a consent form authorizing the experimental use of discarded gametes. These consent forms were scanned and stored in the patient's electronic medical record. The authors did not have access to any identifying personal health information.

Human sperm collection and preparation. Semen samples were collected from five healthy patients between the ages of 30 and 34 undergoing an evaluation for infertility. Each male was a non-smoker with a body mass index $<30 \text{ kg m}^{-2}$. Semen samples were collected by masturbation after 2 days of abstinence. A semen analysis was performed by an experienced andrologist confirming a normal sperm concentration ($>20 \text{ million ml}^{-1}$), normal motility ($>50\%$ motile), and normal morphology using the Kruger strict criteria ($\geq 4\%$ normal forms).

A PureCception gradient solution (PureCception 100% Isotonic Solution, Quinn's Advantage Medium with HEPES, In-Vitro Fertilization Inc.) was prepared to purify each sperm sample and remove somatic cell contaminants. The gradient consisted of two layers of 1 ml of PureCception: 90% and 47% in a 15-ml conical tube.

Two millilitres of the semen sample was placed on top of the gradient. The gradient was centrifuged at 1100 r.p.m. for 20 min and the supernatant removed. One millilitre of sperm washing medium (Quinn's Sperm Washing Medium, In-Vitro Fertilization Inc.) was used to re-suspend the pellet. The sample was then centrifuged at 750 r.p.m. for 10 min. The supernatant was removed. 0.1 ml of the remaining pellet of sperm was transferred to a 1.7 ml SafeSeal Microcentrifuge Tube (Sorenson Bioscience) and placed immediately at minus 80°C .

Human embryo thawing. Excess human embryos created via *in vitro* fertilization for the treatment of infertility were previously donated by patients undergoing assisted reproduction and stored in liquid nitrogen at minus 196°C .

Embryo culture dishes were set up using 60-mm culture dishes (BD Falcon) and eight 30- μl drops of Global embryo culture media (LifeGlobal) plus 15% Plasmanate (Talecris) overlaid with 10 ml of oil (SAGE). Rinse dishes were also set up using 2.5 ml of embryo culture media plus 15% Plasmanate. These dishes were equilibrated overnight at 37°C , 5% CO_2 .

Both cleavage stage embryos and blastocysts were thawed using the Quinn's Advantage Embryo Thaw Kit (SAGE). This kit contains three solutions: 0.5 M sucrose, 0.2 M sucrose and diluent. The straw or the vial containing the cryopreserved embryos was placed in a water bath at 30°C for 2 min. The embryos were then expelled from the straw by removing the heat-sealed end or transferred from the vial using a Pasteur pipette to a clean tissue culture dish on a heated stage. The embryos were located and thawed according to the manufacturer's instructions.

Each embryo was rinsed and then placed in a single drop in the embryo culture dishes described above. Embryos were cultured in a humidified atmosphere at 37°C and 5% CO_2 in air.

Embryo evaluation. Cleavage stage embryos were evaluated 2 to 4 h after the thawing process. Blastocysts were evaluated 18 to 24 h after the thawing process. Embryos were evaluated using a Nikon Eclipse 80i microscope and images of each embryo were obtained at $40\times$ using the Hamilton Thorne Clinical Laser Software. These images were then independently evaluated by two senior embryologists. Survival and quality were determined and an embryo was only included in this study if both embryologists agreed upon viability.

Human embryo collection. Single viable embryos were briefly passed through several rounds of additional defined KSOM media (Millipore) under mineral oil before an Acidic Tyrode's Solution (Sigma) wash to dissolve the zona pellucida, somatic cellular debris, and additional sperm. Single embryos were then rinsed in clean media drops before pooling of the embryos, assessment for the absence of contaminants and snap freezing in minimal volume.

ICM and trophectoderm isolation of human and mouse embryos. All mouse experiments were conducted according to Harvard animal welfare guidelines and approved by the IACUC. Single human and mouse embryos exhibiting a clear ICM were isolated and the zona pellucida carefully removed to preserve the integrity of the blastocoel. They were then positioned using standard micromanipulation equipment (Narishige), oriented with a clear plane available to separate the ICM from the

expanded trophectoderm and dissected using a Hamilton Thorne XYClone laser (Hamilton Thorne Biosciences) with 300 μs pulsing at 100% intensity. Short pulses progressed over the cleavage plane until the ICM and trophectoderm compartments were cleanly separated, at which point both pools were separated for collection in clean microdrops, serially washed and snap frozen in minimal volume. Mouse E.3.5 blastocysts were isolated from hormone-primed C57Bl6/J female mice 4 to 6 weeks of age 3 days after mating with 129S1/SvIMJ males.

Derivation of new human ES cell lines. Human embryo culture and human ES cell derivations were carried out as previously described²⁶. In brief, cleavage stage human embryos were thawed using Quinn's Advantage Thaw Kit (SAGE) and cultured in Global medium (LifeGlobal) supplemented with 15% Plasmanate (Talecris) for 2 to 3 days until the blastocyst stage. For human ES cell derivation, the visible ICM was separated from the blastocyst by exposing the trophectoderm cells to 20- to 30-cell lethal laser pulses from a Hamilton Thorne XYClone laser (Hamilton Thorne Biosciences). The isolated ICM was then plated on a layer of gamma-irradiated mouse embryonic fibroblasts (MEFs) in derivation media consisting of KO-DMEM (Life Technologies), 15% KO-SR (Life Technologies), 2.5% Fetal Bovine Serum (FBS) (Hyclone), 2 mM Glutamax, 1% non-essential amino acids, 50 units per ml penicillin and 50 $\mu\text{g ml}^{-1}$ streptomycin (Life Technologies), 0.055 mM β -mercaptoethanol (Life Technologies), 10 ng ml^{-1} bFGF (Millipore). Ten to twelve days after ICM plating, the embryonic stem cell outgrowth (passage 0) was mechanically dispersed with half of the outgrowth plated onto a new MEF feeder layer for human ES cell line establishment (passage 1), and half used for methylation analysis. The human ES cell colonies that resulted following plating were continuously mechanically passaged. The pluripotency of the lines was confirmed by staining for pluripotency markers and by *in vitro* differentiation into the 3 germ layers. The lines were registered with the Harvard University ESCRO Committee as HUES 71, HUES 72 and HUES 73.

Isolation of mouse E6.5 epiblast and extraembryonic ectoderm. Isolation of E6.5 epiblast and extraembryonic ectoderm was adapted from ref. 27. Hormone-primed C57Bl6/J female mice 4 to 6 weeks of age were euthanized 6 days after mating with 129S1/SvIMJ males. Deciduae were removed from the uterine horn and the full embryo extruded and placed under mineral oil in KSOM media using a pulled glass capillary. Residual maternal contaminants were removed by continuous mouth pipetting, after which the epiblast and extraembryonic ectoderm were bisected using an obliquely cut flame drawn glass capillary and the respective tissues segregated in separate KSOM drops. Visceral endoderm was removed from either epiblast or extraembryonic ectoderm by incubation in 0.5% trypsin, 2.5% pancreatin (Sigma) dissolved in PBS for 20 min at 4°C , after which they were returned to KSOM medium drops and incubated for an additional 5 min at room temperature. Using a glass capillary pulled to a diameter slightly less than that of the embryo, visceral endoderm was removed by rapid aspiration and expulsion. Cleaned epiblast or extraembryonic ectoderm tissue were then serially washed through several additional drops of KSOM before pooling and snap freezing at minimal volume.

Library preparation and sequencing. RRBS libraries were generated as described and sequenced on an Illumina Genome Analyzer II before alignment and analysis⁷. The sequencing reads were aligned to the Human Genome Build 19 (hg19) for human samples and Mouse Genome Build 37 (mm9) for mouse samples using a custom computational pipeline taking into account the strain background for mouse samples. The data set was supplemented with mouse early development methylation profiles from ref. 7 and human fetal somatic methylation profiles from the NIH Epigenomics Roadmap Project. Human ES cell lines H1, H9, HUES64, and HUES6 were used for comparison with our newly derived human ES cell lines. Sample quality was assessed by looking at coverage numbers (that is, number of loci present and coverage of loci) and similarity between biological replicates using Pearson correlation, Euclidean distance, and visual inspection of methylation histograms (see Extended Data Figs 2 and 7).

Estimating methylation levels. The methylation level of each sampled cytosine was estimated as the number of reads reporting a C, divided by the total number of reads reporting a C or T. Single CpG methylation levels were limited to those CpGs that had at least tenfold coverage. For 100-bp tiles, reads for all the CpGs that were covered more than fivefold within the tile were pooled and used to estimate the methylation level as described for single CpGs. The CpG density for a given single CpG is the number of CpGs 50 bp up- and downstream of that CpG. The CpG density for a 100-bp tile is the average of the CpG density for all single CpGs used to estimate methylation level in the tile.

The methylation level reported for a sample is the average methylation level across replicates. A replicate will contribute to the average only if it meets the coverage criteria within the replicate. Technical replicates were averaged before contributing to the sample average.

Genomic features. For mouse, high density CpG promoters (HCPs), intermediate density CpG promoters (ICP), low density CpG promoters (LCPs), CpG island, and DMR annotations were taken from ref. 28. LINE, LTR and SINE annotations

were downloaded from the UCSC (University of California, Santa Cruz) browser (mm9) RepeatMasker tracks. Gene annotations were downloaded from the UCSC browser (mm9) RefSeq track. Promoters (TSSs) are defined as 1 kb up- and downstream of the TSS and are parsed from RefSeq annotation. Promoters for all isoforms are included. Enhancer annotations were taken from Supplementary Table 1 of ref. 29. Corresponding human annotations were downloaded from the UCSC browser for hg19. Human imprinting control regions were taken from ref. 30.

In each case, the methylation level of an individual feature is estimated by pooling read counts for all CpGs within the feature that are covered greater than fivefold, and levels are only reported if a feature contains at least 5 CpGs with such coverage (in contrast to 100-bp tiles where no minimum number of CpGs is required). A tile is annotated as a genomic feature if any portion of the tile overlaps with the feature and thus, may be annotated by more than one feature (for example, the same can be annotated as both a promoter and a gene).

Gene expression analysis. Raw RNA-seq data were downloaded from the gene expression omnibus from accession GSE36552 for human oocytes, zygotes, 2 cell, 8 cell, morulae, late blastocyst, human ES cell passage 0 and human ES cell passage 10, taken from ref. 13. For all samples but morulae and late blastocyst, data for all single cells were pooled before alignment. Morulae and late blastocyst samples were pooled according to their respective embryos. Duplicate reads were removed before alignment. Alignment was performed using TOPHAT against human genome assembly 19 with default settings. Cufflinks was used for quantification and statistical tests of significant change using default settings.

Retrotransposon expression analysis. Alignment was performed using BWA against human genome assembly 19 with default settings. Repeat subfamily FPKM is the sum of the number of reads that align to each repeat element for the subfamily divided by the genome coverage of the subfamily in kilobases and normalized by the total number of reads in the sample. Repeat subfamily FPM is the same as FPKM without normalizing to the subfamily's genome coverage. Samtools was used to find candidate repeat alignments and the CIGAR string was parsed to determine whether the read overlapped with the repeat element. One L1PA6 element (chr 2: 49454725-49460932) contained an unusual tall, short peak of read density and was excluded from analysis.

The L1PA consensus plot for expression was produced by using the consensus coordinates for the element from the UCSC genome browser to align the reads from the element's alignment to the consensus sequence. Read count is the read coverage at each position in the consensus sequence divided by the total number of reads in the sample. Note that we did not use a multiple sequence alignment of the repetitive elements so this is not a fully accurate representation of expression over the consensus. For example, if the element has an insertion compared to the consensus sequence, then reads that overlap the insertion will contribute to read density on the consensus. The genomic sequence density plot was generated in a similar manner.

Orthology between human and mouse. The 46 mammals multiple sequence alignment downloaded from the UCSC browser was used to find orthologous regions from hg19 to mm9. For 100-bp tiles, the methylation for the corresponding region in mouse was used for comparison regardless of the length of the corresponding region. For genomic features, the methylation level of the corresponding region in mouse was used unless the mouse coordinates overlapped a mouse annotation of the same type. In the latter case, the methylation level of the corresponding feature was used instead. The 60 mammals multiple sequence alignment downloaded from the UCSC browser for mm10 was used to find orthologous regions from mouse to human by first translating the mm9 genomic feature coordinates to mm10, and then following the same procedure above.

RRBS selectively enriches for a consistent fraction of CpG dense genomic fragments within a given species, and as such provides genome-scale information versus genome-wide. In human and mouse, the coverage for CGIs is 87.4% and 89.6%, for exons, 11.3% and 8.8%, and for introns, 26.4% and 13.4%, respectively. In general, for both species, most features are captured at similar frequencies, but far more SINEs are captured in human than in mouse. Of the CGIs that are captured by RRBS in human, 82.7% align to the mouse genome, and 53.9% align and share CGI status in mouse. Out of the loci that aligned to the mouse genome, mouse RRBS

captured 93.6% of shared CGIs and 72.9% of aligned regions not annotated as CGIs in mouse. 85.2% of RRBS-captured human exons align to mouse exons and, of these, 52.7% are captured by mouse RRBS. For introns, 86.1% of human introns align to mouse introns and, of these, 34.2% are captured in mouse.

Derivation sliders used to compare human ES cells to mouse pre and post implantation pluripotent tissues. Regions with a methylation difference >0.1 between mouse ICM and mouse epiblast were used to assess the similarity of a sample to either of these tissues. For each sample, a region scores as mouse ICM if its methylation level is more similar to mouse ICM than mouse epiblast and vice versa. Regions with methylation values that are equidistant from mouse ICM and mouse epiblast are excluded. If the slider is viewed as going from 0 (mouse ICM) to 1 (mouse epiblast), then the position of a sample is simply the proportion of regions that scored mouse ICM over the total number of regions that contribute a score.

Clustering and feature enrichment. One hundred base pair tiles were clustered using k-means clustering. Clusters were designated as hypermethylated (≥ 0.5) or hypomethylated (<0.5) in sperm according to the cluster centre. A tile and a feature were designated as overlapping if there was an overlap of 1 bp or more between them. Feature enrichment scores are the negative log of the *P* value calculated using the hypergeometric distribution.

Identification of gametic differentially methylated regions in human. Regions that have low methylation in sperm and some methylation in the early embryo likely represent instances of maternal methylation if the assumption that there is little to no *de novo* methylation over the cleavage divisions is true. A region is considered a maternally methylated differentially methylated region if: first, it is significant after a two sample *t*-test between sperm and blastocyst with equal variance after correction for multiple hypothesis testing (*q* value <0.05 using the Storey method, ref. 31); second, it has a methylation difference ≥ 0.2 higher in blastocyst than in sperm using sample means; and third, it has a mean methylation level ≥ 0.2 for 8 cell. These criteria were applied to both 100-bp tiles and to CGIs.

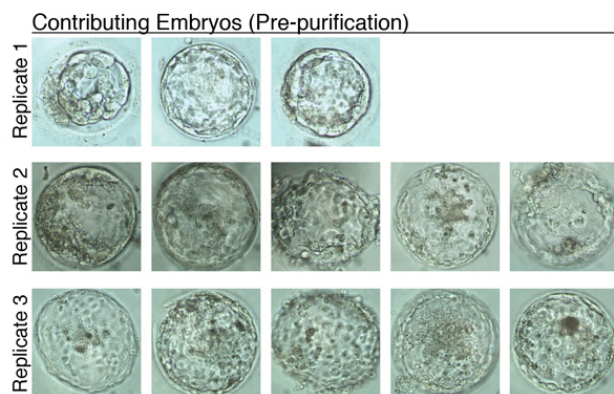
Identification of human SNPs. SNPs in human were downloaded from the 1000 Genomes Project (<ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/release/20110521/>). SNPs that are not trackable by RRBS (C/T or A/G) and positions that are not covered in an *in silico* digest of hg19, that is, covered by RRBS, were removed. The ratio (reference count/(reference count + alternative count)) was calculated for each SNP in the single human blastocyst samples and SNPs with ratios <0.2 or ratios >0.8 were removed since they likely represent homozygous alleles in the sample. We used the resulting genotypes to facilitate parent-of-origin methylation tracking.

Parent-of-origin methylation tracking. Reads were segregated into either the reference or alternative allele, and CpG methylation levels were called in the same manner described above. SNP normalized methylation values (Extended Data Fig. 7) are the average of the methylation values derived from each haplotype.

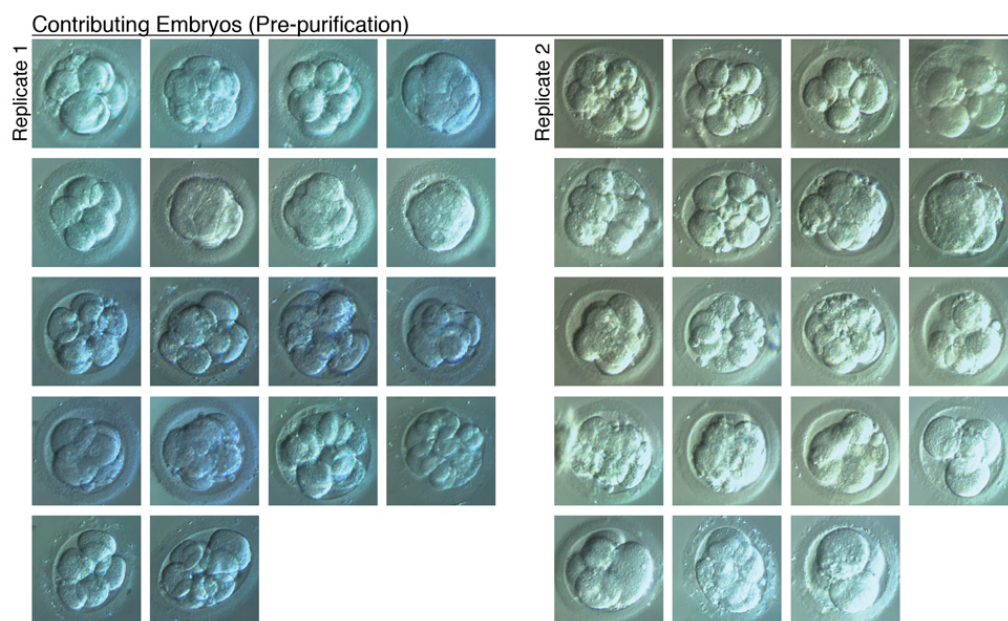
L1PA Sequence Alignment. All LINE L1HS and L1PA2-7 elements $\geq 5,500$ bp were taken from the human genome and aligned using Muscle with the following parameters: Muscle – Maxiters 2 – Diags. Methylation levels are from full-length elements that are captured by RRBS. The ~ 130 bp insert sequences common to elements L1PA3b and older were similarly aligned for Extended Data Fig. 10d.

26. Chen, A. E. *et al.* Optimal timing of inner cell mass isolation increases the efficiency of human embryonic stem cell derivation and allows generation of sibling cell lines. *Cell Stem Cell* **4**, 103–106 (2009).
27. Chenoweth, J. G. & Tesar, P. J. Isolation and maintenance of mouse epiblast stem cells. *Methods Mol. Biol.* **636**, 25–44 (2010).
28. Meissner, A. *et al.* Genome-scale DNA methylation maps of pluripotent and differentiated cells. *Nature* **454**, 766–770 10.1038/nature07107 (2008).
29. Creighton, M. P. *et al.* Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl Acad. Sci. USA* **107**, 21931–21936 (2010).
30. Woodfine, K., Huddleston, J. E. & Murrell, A. Quantitative analysis of DNA methylation at all human imprinted regions reveals preservation of epigenetic stability in adult somatic tissue. *Epigenet. Chromatin* **4**, 1 (2011).
31. Storey, J. D. & Tibshirani, R. Statistical significance for genomewide studies. *Proc. Natl Acad. Sci. USA* **100**, 9440–9445 (2003).

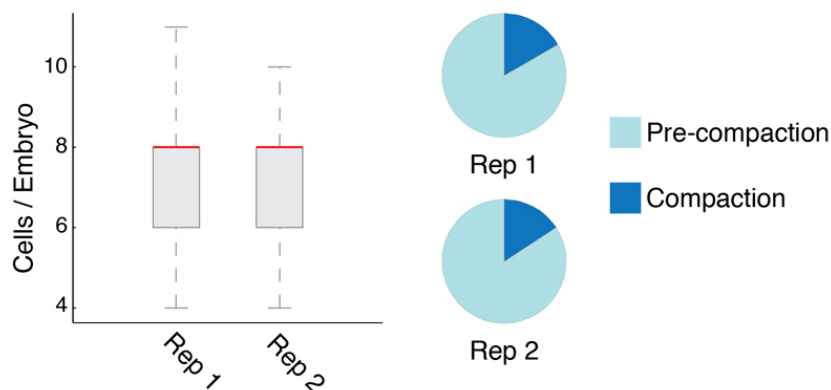
a



b

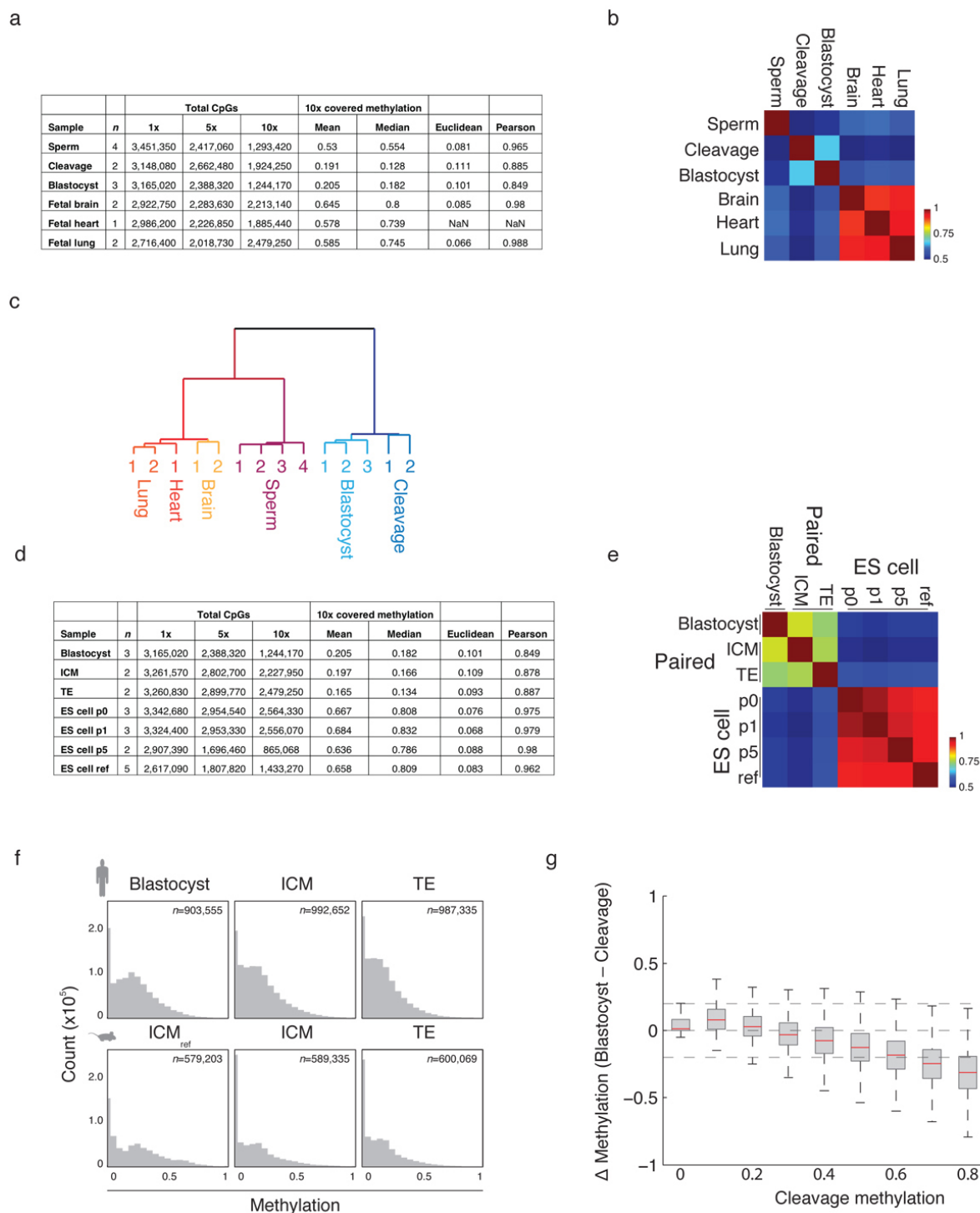


c



Extended Data Figure 1 | Isolation of human preimplantation embryos for DNA methylation mapping. a, Three replicates of D6 embryos, ranging in inputs from three to five embryos, were thawed, screened for proper morphology, independently isolated from the zona pellucida and pooled before RRBS profiling. Embryos are displayed before purification. b, Two replicates of 18 and 19 human D3 cleavage stage embryos were thawed, screened for proper morphology, assessed for embryonic stage and cell number, and

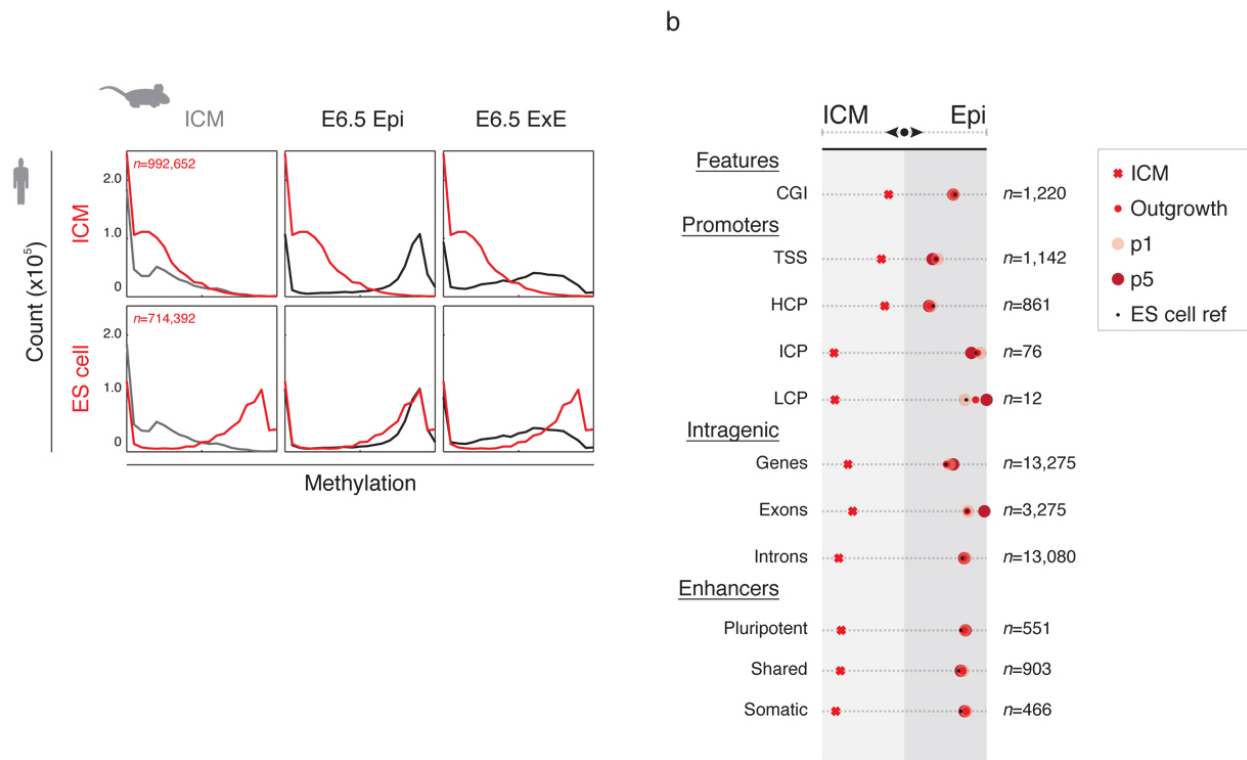
purified from the zona pellucida before RRBS profiling. Embryos are displayed before purification. c, Cell numbers from thawed cleavage stage embryos ranged from 4 to 11 cells per embryo with a median of 8 (± 1.6 standard deviation) cells. Within each replicate, only three embryos demonstrated onset of compaction at the time of collection. Red line signifies the median, boxes and whiskers the 25th and 75th, and 2.5th and 97.5th percentiles, respectively.



Extended Data Figure 2 | Assembly of a genome-scale DNA methylation time series through human early development and over ES cell derivation.

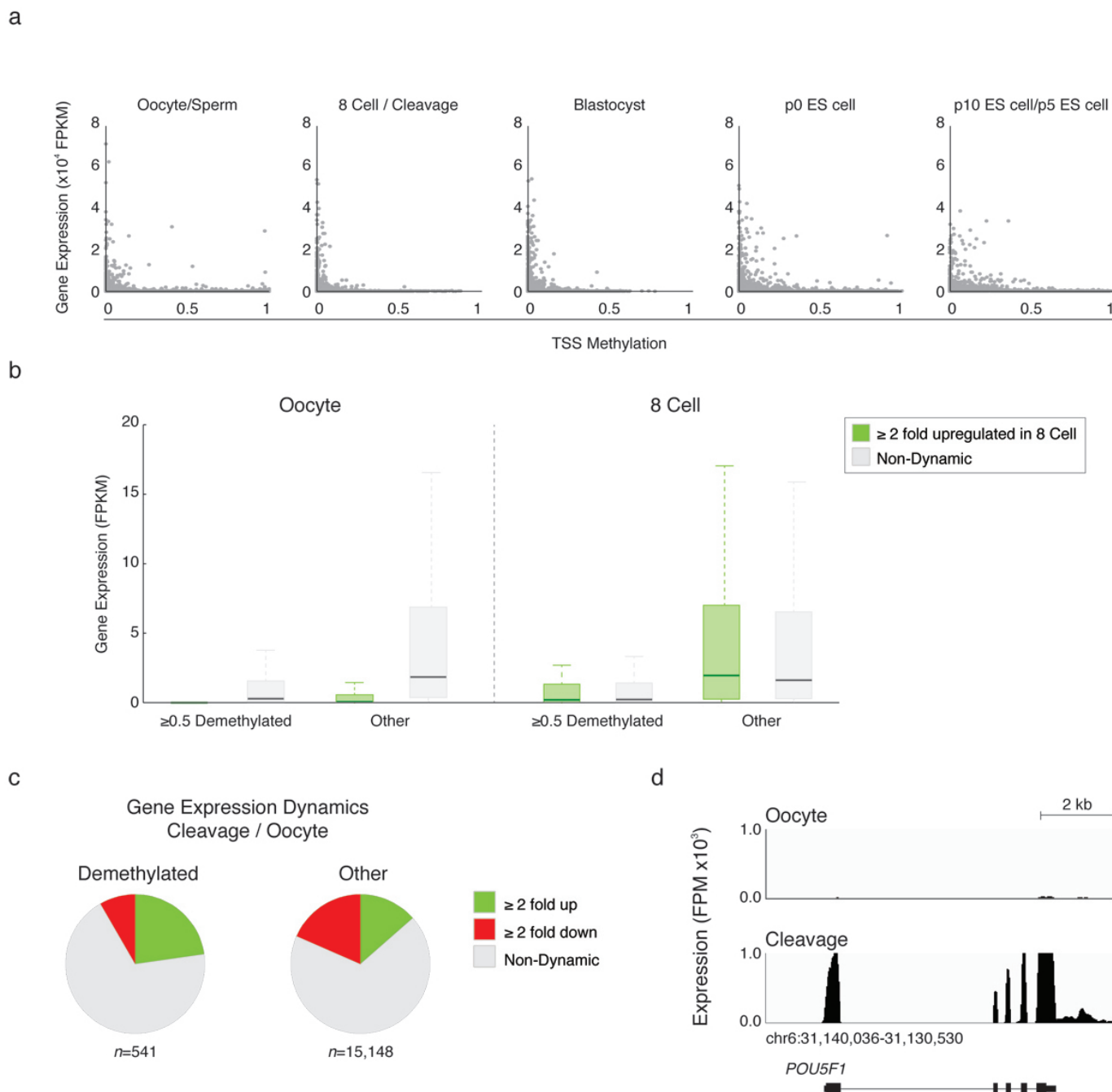
a, Summary of RRBS libraries generated, with number of biological replicates (*n*), number CpGs captured at 1×, 5× and 10×, mean and median methylation values for 100-bp tiles estimated from CpGs covered at ≥5×, and mean Euclidean distance and Pearson correlation across biological replicates for these tiles. **b**, Pearson correlation matrix for sperm, early embryonic and fetal tissue samples. **c**, Clustering of gametic, somatic and preimplantation methylation profiles segregate according to their global DNA methylation landscape, with sperm and fetal tissue forming a somatic methylation cluster that contrasts the unique epigenetic landscape present in preimplantation embryos. **d**, Summary of RRBS libraries generated for ES cell derivation, with number of biological replicates (*n*), number CpGs captured at 1×, 5× and 10×, mean and median methylation values for 100-bp tiles estimated from CpGs covered at ≥5×, and mean Euclidean distance and Pearson correlation across biological replicates

for these tiles. 'ES cell ref' refers to a reference collection of previously assayed ES cell lines as part of the NIH Roadmap Epigenomics Project (Methods). Human ICM and trophectoderm (TE) were isolated through laser-assisted microdissection. **e**, Pearson correlation matrix for human samples used to model ES cell derivation. A consistent signature is rapidly acquired by the outgrowth stage (p0) and stably maintained over additional passages. **f**, Methylation histograms for 100-bp tiles for human blastocysts and dissected ICM and TE tissue show minimal global difference, which is also observed when comparing previously assayed, immunosurgically purified mouse ICM to mechanically dissected ICM and TE. **g**, Boxplots of the change in methylation (Δ methylation) for 100-bp tiles from cleavage to the blastocyst stage show passive demethylation of DNA methylation, particularly for regions that exhibit the highest methylation levels at this stage. The red line signifies the median, boxes and whiskers the 25th and 75th, and 2.5th and 97.5th percentiles, respectively.



Extended Data Figure 3 | Comparison of human ES cell derivation to *in vivo* mouse pluripotent tissues. **a**, Global methylation histograms of 100-bp tiles for human ICM and p5 ES cells (rows) compared against mouse preimplantation and postimplantation embryos (ICM, E6.5 Epi), as well as with extraembryonic ectoderm (ExE) (columns), demonstrate the rapid acquisition of an epiblast-like, somatic methylation pattern upon ES cell derivation in human. **b**, Regions that discriminate mouse ICM from E6.5 epiblast were used to assign human ICM and ES cells to an equivalent *in vivo* pluripotent state for

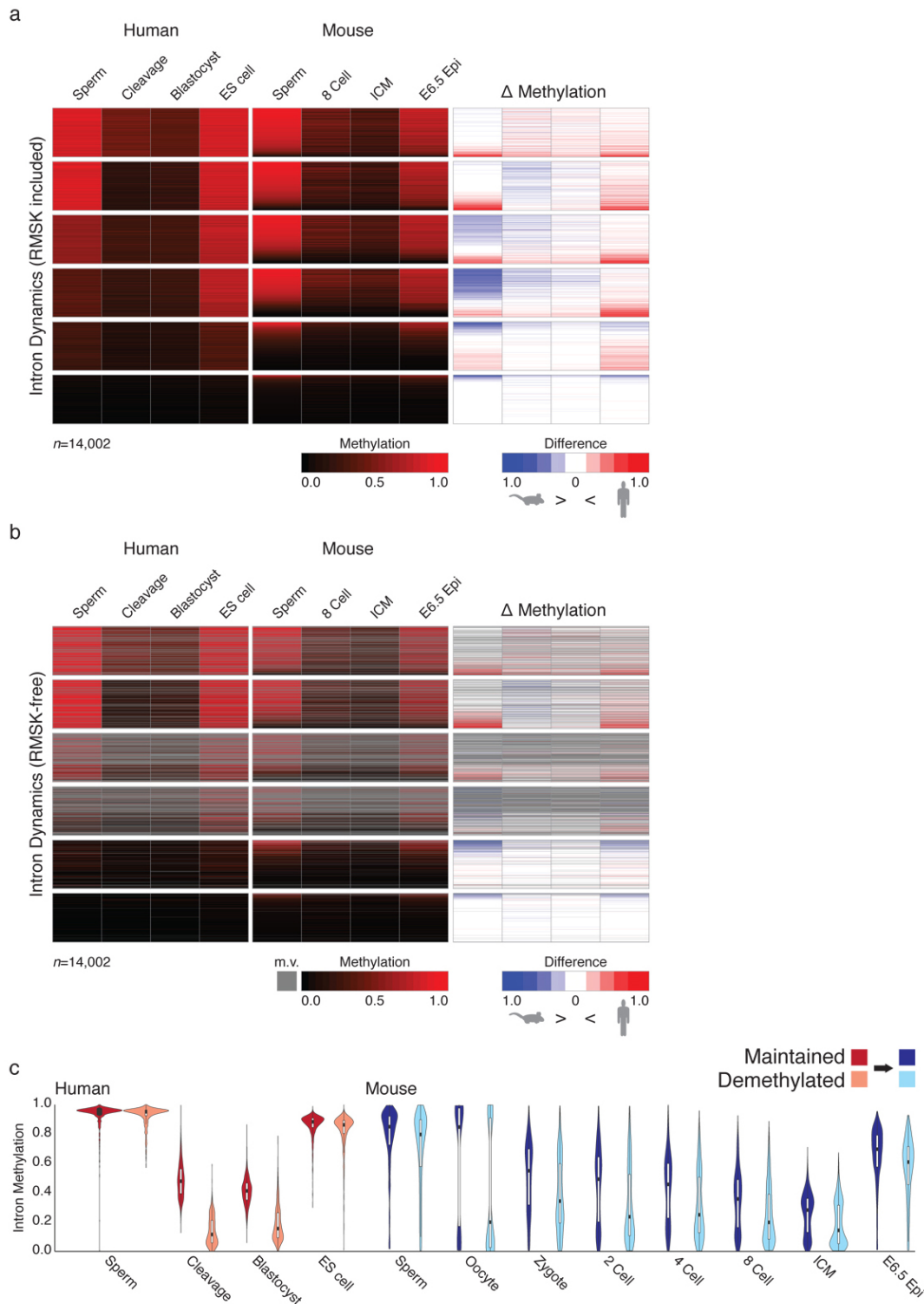
orthologous features. The position along the axis from preimplantation (ICM) to postimplantation (Epi) pluripotency represents the proportion of regions in a set that resemble one state versus the other. For all feature sets, human ES cells rapidly establish an epiblast identity, maintaining this signature from the outgrowth stage over ensuing passages. 'ES cell ref' refers to a reference collection of previously assayed ES cell lines as part of the NIH Roadmap Epigenomics Project (Methods).



Extended Data Figure 4 | Inverse correlation between expression and promoter methylation is retained during human preimplantation.

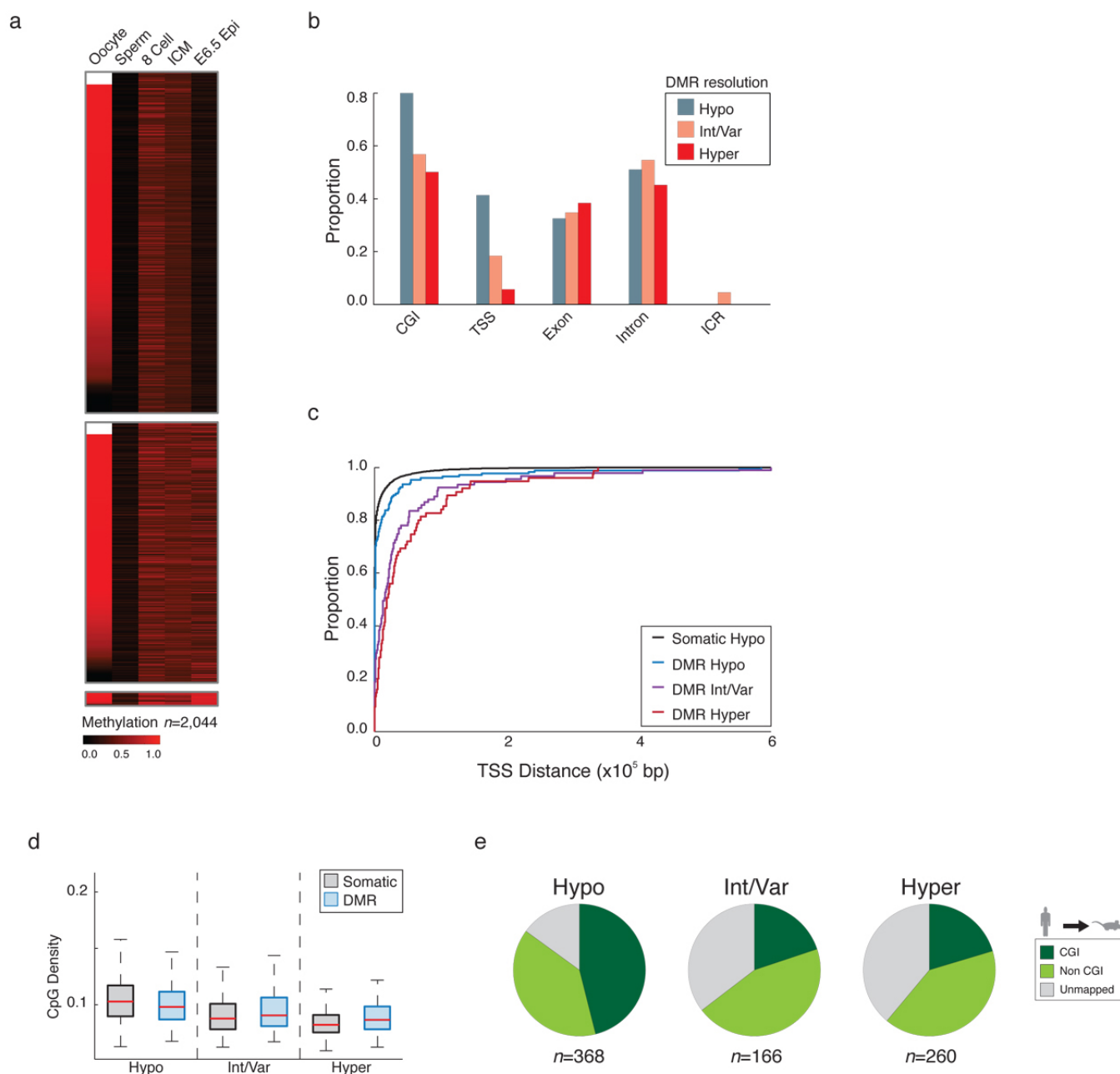
a, Scatterplots of oocyte, preimplantation embryo and ES cell derivation gene expression compared to promoter methylation display a canonical negative correlation, even during preimplantation where the range of promoter methylation values is contracted by global hypomethylation. **b**, Box plots of gene expression values for genes significantly upregulated by ≥ 2 -fold from oocyte to 8 cell compared to non-dynamic genes and categorized by promoter methylation dynamics. Genes that are both demethylated and upregulated are associated with induction from a silenced state, while those that are demethylated but not upregulated display only basal level transcription that is significantly lower than observed in promoters that are not demethylated. Bold

line signifies the median, boxes and whiskers the 25th and 75th, and 2.5th and 97.5th percentiles, respectively. **c**, Gene expression dynamics following fertilization for hypermethylated sperm promoters demethylated ≥ 0.5 by the cleavage stage compared to the rest of promoters (Other). 123 of 541 (22%) demethylated promoters demonstrate significant upregulation (≥ 2 -fold) compared to only 13.6% of other promoters. Moreover, the ratio of upregulated to downregulated genes in the demethylated set substantially favours zygotic activation, while other promoters include a higher proportion of downregulated maternal transcripts (odds ratio = 1.877, $P = 1.344 \times 10^{-8}$, hypergeometric test). **d**, RNA-seq track of the pluripotency promoting, zygotically induced gene *POU5F1*, whose promoter is demethylated from 0.59 in sperm to 0.02 in cleavage, concurrent with its transcriptional induction.



Extended Data Figure 5 | Local retention of DNA methylation is similar for introns over human and mouse preimplantation. **a**, Introns are clustered according to their dynamics in human and the methylation of orthologous regions are tracked in mouse. Divergence is predominantly restricted to intermediately methylated features in human sperm that are generally hypermethylated in mouse. The Δ methylation heatmap displays the difference in methylation values between equivalent preimplantation time points, with ES cells in human serving as a proxy for comparison to the E6.5 epiblast in mouse. Deviation is most apparent for intermediately methylated human sperm introns, where they are less methylated than in mouse sperm. RMSK included, repeat masker annotated regions included. n shows the number of introns available for cross species comparison. **b**, When repetitive elements are

removed from the calculation of intron methylation, the apparent divergence between mouse and human values in sperm is diminished. Methylation and Δ methylation heatmaps are as in **a**. Gray denotes missing values (m.v.) where estimates for intronic methylation were exclusively derived from repetitive elements. RMSK-free: repeat masker annotated regions excluded. **c**, Violin plots of the two main dynamics (maintained versus demethylated, top two clusters in **b**) for sperm hypermethylated introns over human and mouse preimplantation after repetitive elements are removed. As observed for orthologous exons, regions that retain high methylation throughout human preimplantation are conserved, hypermethylated in both mouse gametes, and display maintained regulation as early as the zygote stage.



Extended Data Figure 6 | Genomic characterization of transient maternally contributed imprint-like regions.

a, Heatmap of 100-bp tiles in mouse preimplantation identified using the same criteria as applied to human (Methods). This criteria, which assumes limited *de novo* methylation, identifies 2,044 tiles in mouse where methylation is ≥ 0.2 in both 8 cell and the ICM, there is ≥ 0.2 methylation difference between the ICM and sperm, and this difference is significant via *t*-test, (*q* value < 0.05). 89% of those tiles that are captured in the mouse oocyte are monoallelically inherited and show significant differences between the gametes by *t*-test, providing an empirical upper bound on the false discovery rate for this strategy when applied to human of ≤ 0.11 , assuming the underlying principles of imprint regulation are the same as in mouse during this developmental stage.

b, The proportion of 100-bp tiles, classified according to their resolution in ES cells, for each genomic feature presented in Fig. 3a.

c, Cumulative density function (CDF) plot of the distance to the nearest annotated TSS for CGI DMRs that resolve to hypomethylation, intermediate or variable methylation, or hypermethylation. There is a discrepancy in genomic location between those that resolve to hypomethylation, of which a sizable fraction are in the TSS, and those that do

not, which are generally enriched further downstream.

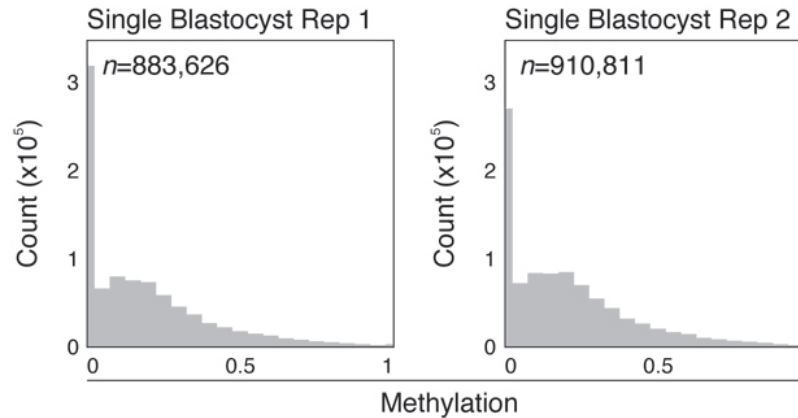
d, Boxplots of CpG density for CGI DMRs that resolve to hypomethylation, intermediate or variable methylation, or hypermethylation paired with comparable non-DMR CGIs (Somatic). Those resolving to hypomethylation have higher CpG densities than those that resolve to intermediate or variable, or hypermethylation, but have slightly lower CpG density than non-DMR, constitutively hypomethylated CGIs. Alternatively, while CGIs that resolve to hypermethylation show a lower CpG density than other DMRs, they show higher density than non-DMR hypermethylated islands, suggesting some level of protection against deamination as an attribute of their uniquely hypomethylated status in the male germline.

e, Pie charts of cross species alignment and CGI status of human CGI DMRs into mouse. Those that resolve to hypomethylation are more often conserved in mouse and more frequently retain their CGI status, whereas those resolving to hypermethylation are less conserved. Moreover, intermediate/variable and hypermethylation-resolving regions that do align are less frequently retained as CGIs. 368, 166 and 260 CGIs comprise the hypo, intermediate or variable, and hyper methylation sets, respectively.

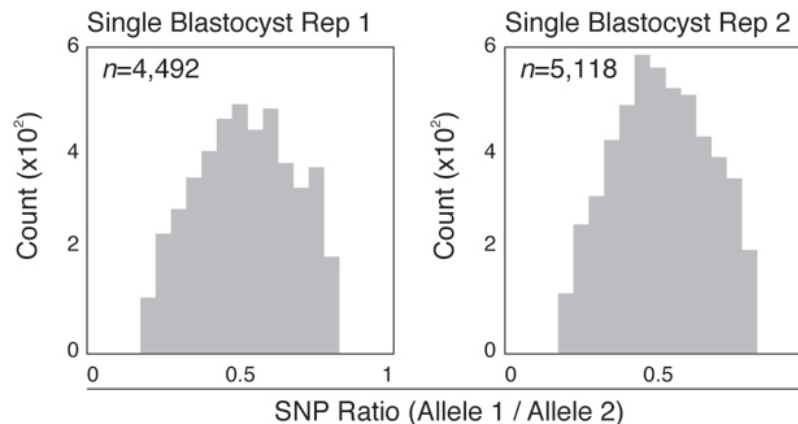
a

Single Blastocyst	Total CpGs			10x covered methylation		Blast comparison	
	1x	5x	10x	Mean	Median	Euclidean	Pearson
Replicate 1	3,123,020	2,745,070	2,262,670	0.17	0.111	0.109	0.65
Replicate 2	3,147,640	2,820,190	2,364,930	0.19	0.143	0.1	0.691

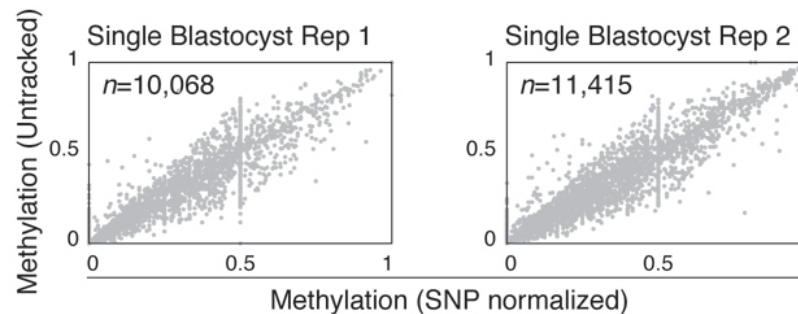
b



c

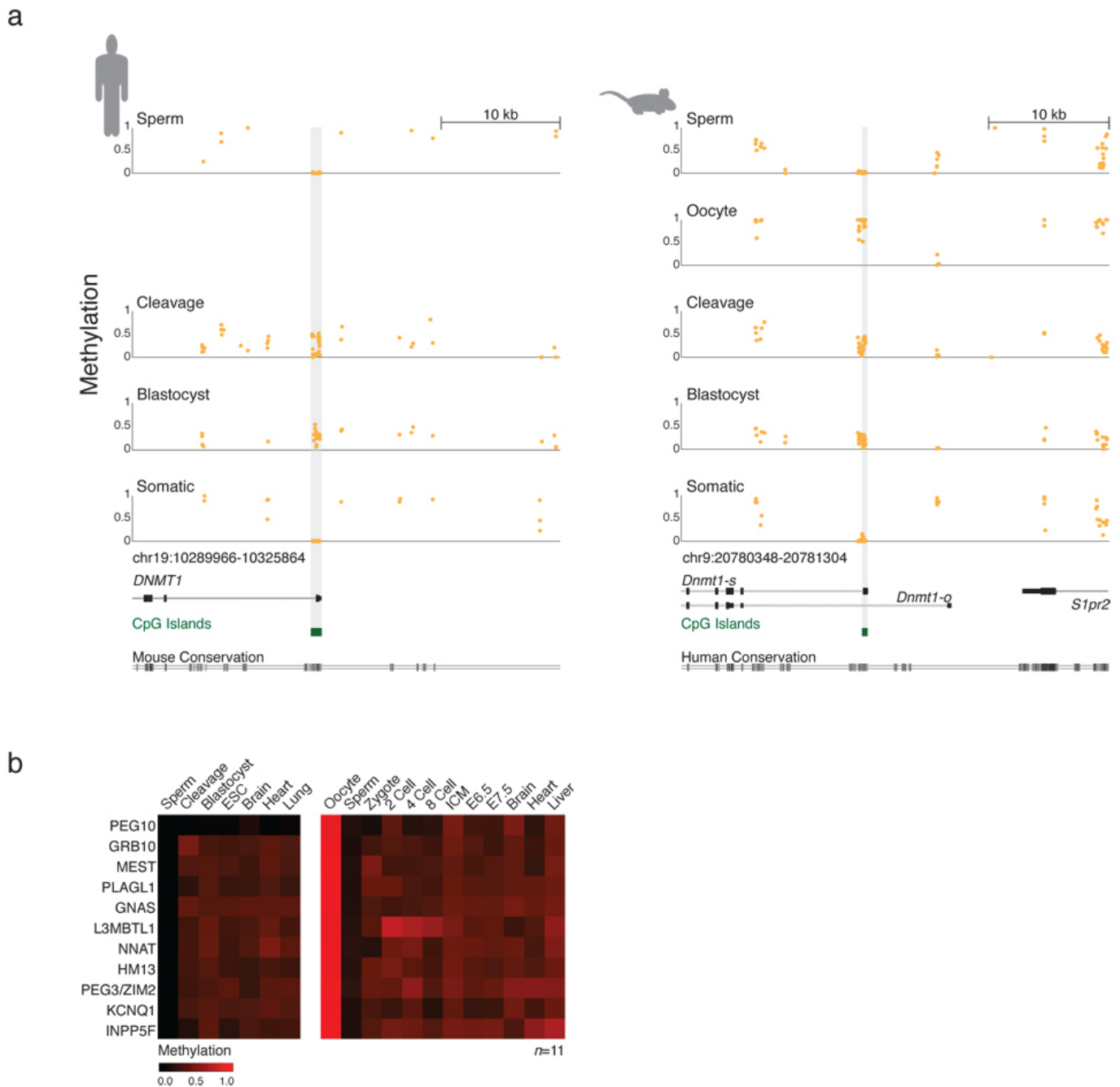


d



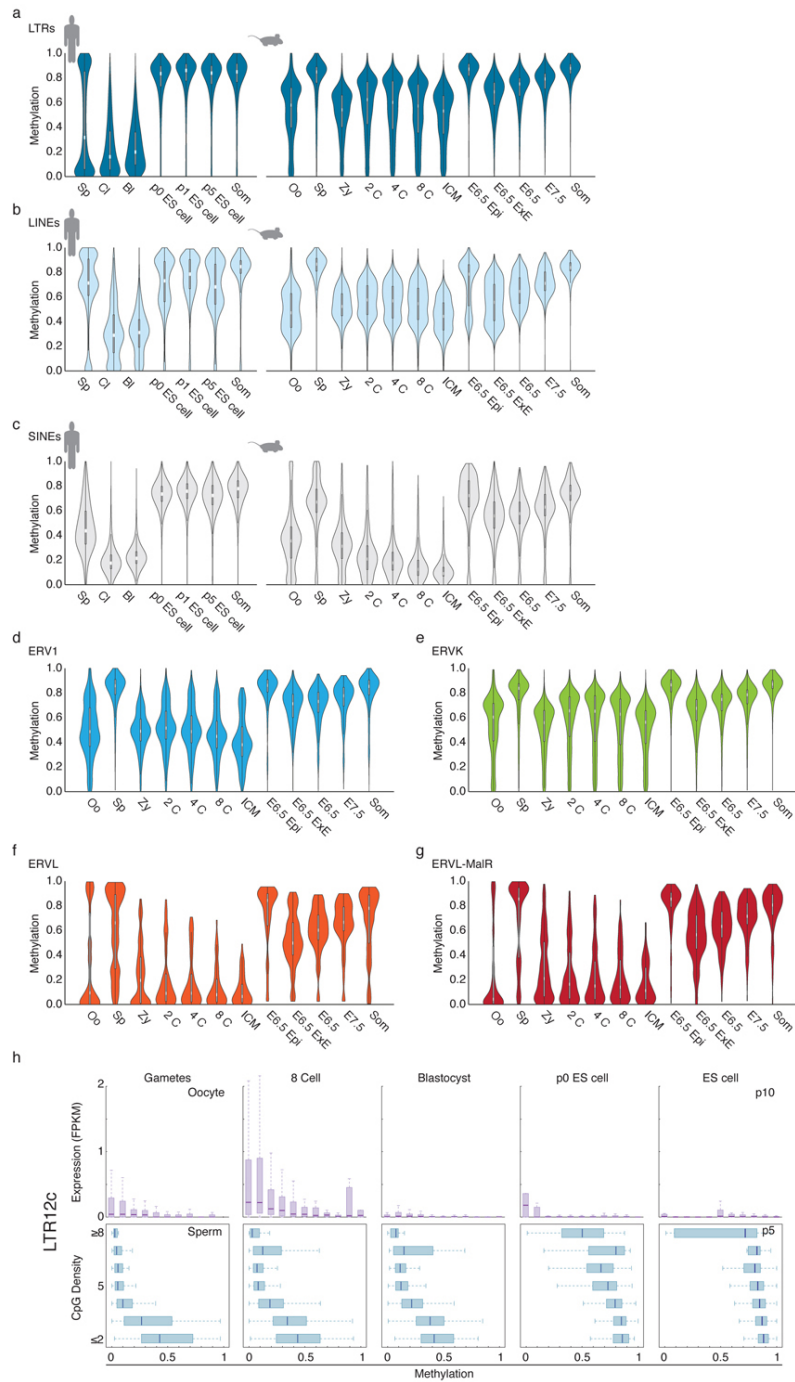
Extended Data Figure 7 | Generation of single blastocyst libraries confirm the monoallelic behaviour of putative maternal DMRs. **a**, Summary of two single blastocyst RRBS libraries. Number of CpGs captured at 1 \times , 5 \times and 10 \times , mean and median methylation values for 100-bp tiles estimated from CpGs covered at $\geq 5\times$, and mean Euclidean distance and Pearson correlation when single blastocyst replicates are compared to the pooled blastocyst time point. **b**, Histograms of DNA methylation for 100-bp regions captured for each single

blastocyst replicate. **c**, The ratio of reference allele to alternative allele for single nucleotide polymorphisms (SNPs) called as heterozygous in each blastocyst replicate. **d**, For the 4,492 and 5,118 SNPs that were considered as heterozygous within each single blastocyst, 10,068 and 11,415 single CpGs could be assigned to alleles. Scatterplots depict untracked methylation values for these CpGs against their normalized methylation values, which are the average of their monoallelic methylation states.



Extended Data Figure 8 | The somatic promoter of *DNMT1* is maternally methylated in human and mouse. **a**, Plots of single CpG methylation for *DNMT1*, including a CGI over the somatic promoter that behaves as a transient, preimplantation-specific DMR in both human and mouse. In mouse, hypermethylation of this island corresponds to its transcriptional readthrough

and exclusion as part of an oocyte-specific isoform (*Dnmt1-o*) that is not annotated in human. Annotated CGIs and species conservation tracks are included for reference below. **b**, Heatmap of orthologous ICR dynamics over human and mouse preimplantation. Of those that map between species and are captured by RRBS, all but one (*PEG10*) behave identically.



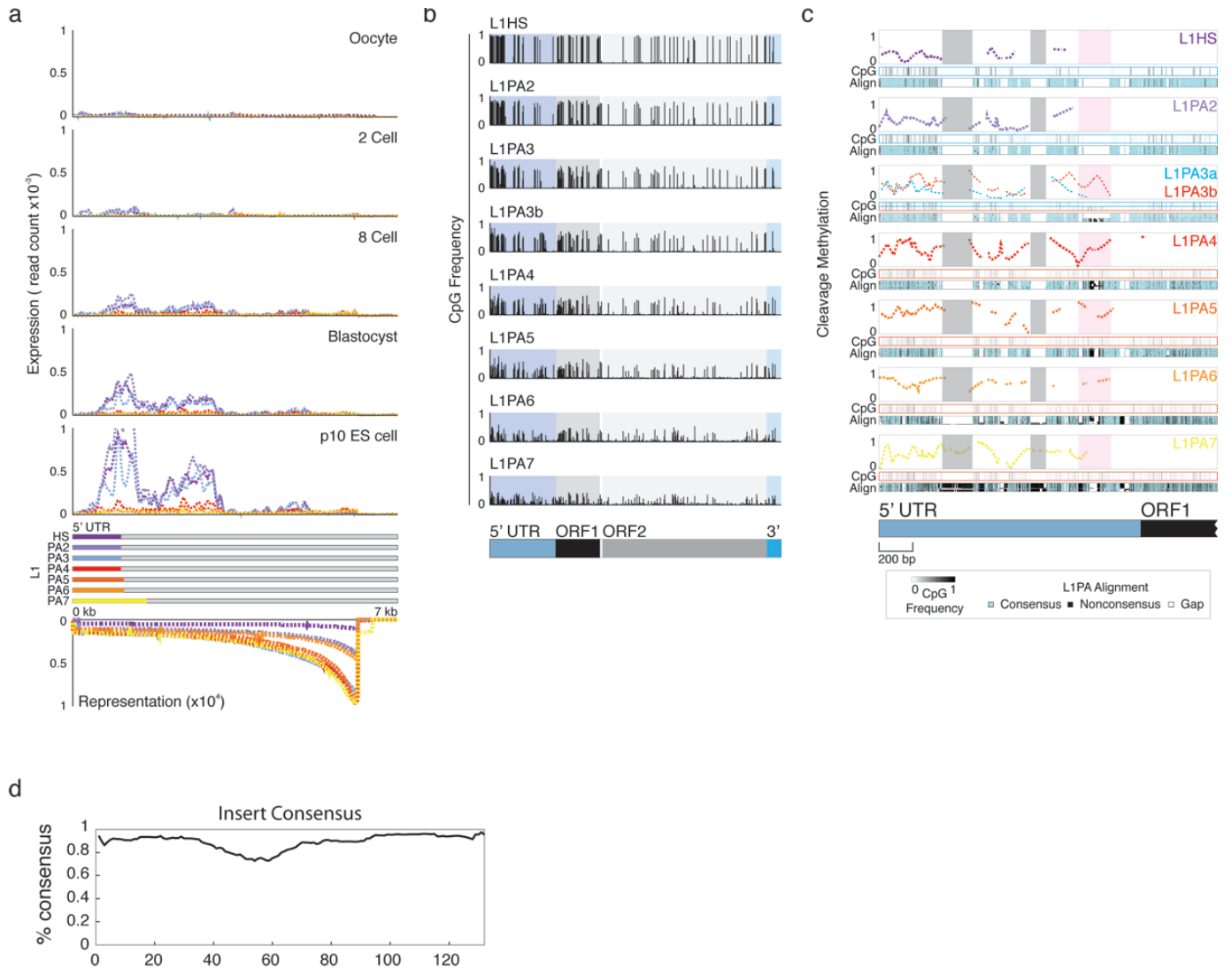
Extended Data Figure 9 | Repetitive element regulation during human and mouse preimplantation.

a, Violin plots for LTRs over human and mouse development. In human, LTRs demonstrate a bimodal distribution in sperm. Hypermethylated LTRs display a range of demethylation in the early embryo that reflects the dynamics of subfamilies. Upon ES cell derivation, and within fetal tissues, LTRs become stably hypermethylated. Alternatively, during mouse preimplantation, LTRs are consistently hypermethylated in sperm and generally retain methylation over preimplantation. E6.5 Epi and E6.5 ExE refer to dissected epiblast and extraembryonic tissue from E6.5 embryos. **b**, Violin plots for LINEs over human and mouse development. In human sperm, LINEs are unstably hypermethylated, with discrete populations methylated with a mean of ~ 0.75 , ≥ 0.9 , and a small subpopulation showing gametic escape from high methylation. Alternatively, LINEs are indiscriminately hypermethylated in mouse sperm. In both species, several populations of elements demonstrate different extents of demethylation during preimplantation, including many that retain higher levels in cleavage and only minor, passive depletion into blastocyst. Upon human ES cell derivation or during mouse implantation, elements are generally remethylated, though only partially for those elements that are demethylated after fertilization. Hypermethylation is complete in fetal tissue. In human, these discrete dynamics can be attributed to the unstable methylation for L1HS-L1PA3a subfamilies while, in mouse, subsets of L1Md_Tf and L1Md_Gf subfamilies are similarly demethylated and elements of the independently emerging L1Md_A lineage remain largely methylated. **c**, Violin plots for SINEs highlight intermediate methylation in sperm in both species, though more so for humans. After fertilization, SINE methylation rapidly diminishes to near complete hypomethylation over preimplantation, similar to what is observed for intergenic sequence, before complete hypermethylation during ES cell derivation in human or in postimplantation mouse E6.5 embryos. Taken globally, SINEs appear to be uniformly regulated regardless of subfamily, though differences in regulatory status for specific

SINE elements may be reflected by their surrounding genomic context.

Unfortunately, such inferences require higher genomic resolution than is currently available to distinguish the dynamics of specific integrations.

d–g, Violin plots of the four major LTR families present in mouse over the complete preimplantation timeline. ERV1 elements (**d**) are hypermethylated in sperm and display a range of demethylation following fertilization and prompt remethylation upon implantation. In mouse, ERVK elements (**e**) are emergent and largely consist of the dominating, constitutively hypermethylated IAP subfamilies. ERVL and MalR (ERVL-MalR) elements (**f** and **g**), the evolutionarily oldest mammalian LTRs, are hypermethylated in sperm and rapidly demethylated after fertilization, frequently in association with their rapid zygotic induction. **h**, Distribution (as boxplots) of per element expression and CpG density at different methylation levels for LTR12c demonstrates negative correlation between methylation and expression. On average, LTR12C is hypomethylated in sperm and the early embryo, but demonstrates a consistent range of values at the level of single elements, with least methylated elements contributing the most to LTR12c expression. The CpG density of these elements corresponds to their degree of hypomethylation, suggesting that escape from *de novo* methylation during spermatogenesis and preimplantation is maintained for specific elements over generations. Once targeted, element expression is apparently restricted and its CpG density decays correspondingly. During ES cell derivation, the kinetics of LTR12c methylation is more rapid for those of lower CpG density, as evident from p0 to p5 in the ES cell lines. DNA methylation in the early embryo is therefore not exclusive to the regulation of different ERV1 subfamilies, but also affects the contribution of single elements to the broader transcriptional pattern. Bold line signifies the median, boxes and whiskers the 25th and 75th, and 2.5th and 97.5th percentiles, respectively. Expression is calculated as the number of fragments per million that align to a given element divided by its length in kb (FPKM).



Extended Data Figure 10 | L1PA subfamily dynamics during human early development. **a**, Expression composite averaged by genomic representation for L1HS through L1PA7 from oocyte through preimplantation and ES cell derivation. Dynamic expression within the L1PA phylogeny is restricted to the same subfamilies that are demethylated by cleavage. The position of each respective 5' UTR, the functional promoter for LINEs, is highlighted in the legend. Beneath these composites is the genomic representation to the full-length consensus for each annotated L1PA subfamily, which demonstrates relative equivalence of 5' UTR representation across different subfamilies, but an increasing proportion of truncated 3' fragments with subfamily age (Methods). **b**, The frequency of CpGs within aligned L1PA subfamilies, including 5' UTR, ORF1 and ORF2, and 3' UTR. CpGs are primarily enriched within the 5' UTR promoter and become progressively CpG depleted with

element age. **c**, Complete composite plot of cleavage stage methylation values across aligned 5' UTRs from L1HS through L1PA7 as in Fig. 5d. The multiple sequence alignment for each subfamily to the assembled consensus is visualized below each composite, with blue corresponding to conservation, black to divergence, and white to gaps or deletions. The x axis represents position along the 5' UTR and a portion of ORF1 for the L1HS consensus. CpG Frequency describes the level of conservation for individual CpGs found within single elements to the consensus. The ~130-bp sequence present from L1PA7 to L1PA3b and absent from L1PA3a to L1HS is highlighted in pink, while two older sequences specific to L1PA7 are highlighted in grey. **d**, Percent identity to the consensus for the extracted ~130-bp insert sequence in elements from L1PA7 through L1PA3b. Mean nucleotide identity to the consensus is 85%, with a median of 89%.

Targeting transcription regulation in cancer with a covalent CDK7 inhibitor

Nicholas Kwiatkowski^{1,2,3*}, Tinghu Zhang^{1,2*}, Peter B. Rahl³, Brian J. Abraham³, Jessica Reddy^{3,4}, Scott B. Ficarro^{1,2,5}, Anahita Dastur⁶, Arnaud Amzallag^{6,7}, Sridhar Ramaswamy^{6,7}, Bethany Tesar^{8,9}, Catherine E. Jenkins¹⁰, Nancy M. Hannett³, Douglas McMillin^{8,9}, Takaomi Sanda^{11,12}, Taebo Sim¹³, Nam Doo Kim¹⁴, Thomas Look^{11,15}, Constantine S. Mitsiades^{8,9}, Andrew P. Weng¹⁰, Jennifer R. Brown^{8,9}, Cyril H. Benes⁶, Jarrod A. Marto^{1,2,5}, Richard A. Young^{3,4} & Nathanael S. Gray^{1,2}

Tumour oncogenes include transcription factors that co-opt the general transcriptional machinery to sustain the oncogenic state¹, but direct pharmacological inhibition of transcription factors has so far proven difficult². However, the transcriptional machinery contains various enzymatic cofactors that can be targeted for the development of new therapeutic candidates³, including cyclin-dependent kinases (CDKs)⁴. Here we present the discovery and characterization of a covalent CDK7 inhibitor, THZ1, which has the unprecedented ability to target a remote cysteine residue located outside of the canonical kinase domain, providing an unanticipated means of achieving selectivity for CDK7. Cancer cell-line profiling indicates that a subset of cancer cell lines, including human T-cell acute lymphoblastic leukaemia (T-ALL), have exceptional sensitivity to THZ1. Genome-wide analysis in Jurkat T-ALL cells shows that THZ1 disproportionately affects transcription of *RUNX1* and suggests that sensitivity to THZ1 may be due to vulnerability conferred by the *RUNX1* super-enhancer and the key role of *RUNX1* in the core transcriptional regulatory circuitry of these tumour cells. Pharmacological modulation of CDK7 kinase activity may thus provide an approach to identify and treat tumour types that are dependent on transcription for maintenance of the oncogenic state.

In an effort to discover new inhibitors of kinases that regulate gene transcription, we performed cell-based screening and kinase selectivity profiling of a library of known and novel ATP-site-directed kinase inhibitors (see Supplementary Table 1 for known CDK7 inhibitors). We identified THZ1 (Fig. 1a), a phenylaminopyrimidine bearing a potentially cysteine-reactive acrylamide moiety, as a low nanomolar inhibitor of cell proliferation and biochemical CDK7 activity (Fig. 1b, c). To investigate the functional relevance of the acrylamide moiety, we prepared a non-cysteine reactive analogue, THZ1-R, which showed comparatively less activity on CDK7 and reduced antiproliferative potency (Fig. 1b, c). KinNav profiling⁵, which measures the ability of a compound to block nucleotide-dependent enzymes from biotinylation with a reactive desthiobiotin-ATP probe, established CDK7 as the primary intracellular target of THZ1, but not of THZ1-R (Supplementary Table 2). Kinome-wide profiling identified additional kinase targets of THZ1; however, we confirmed CDK7 as the only target showing time-dependent inhibition, which is suggestive of covalent binding (Extended Data Fig. 1a–c and Supplementary Table 3).

As no covalent inhibitors of CDKs have been reported, we next focused our studies on the mechanism by which THZ1 could achieve covalent inhibition of CDK7. We first incubated a recombinant CDK7–cyclin H–MAT1 (MAT1 also known as MNAT1) trimeric complex with a biotinylated

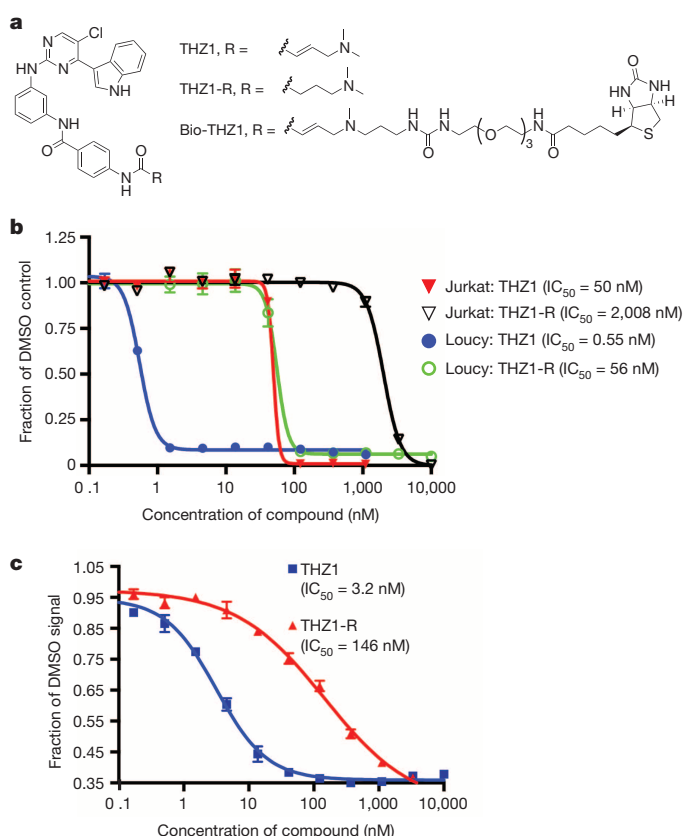


Figure 1 | Cell-based screening and kinome profiling identifies phenylamino-pyrimidines as a potential CDK7 scaffold. **a**, Compound structures of THZ1, THZ1-R and bio-THZ1. **b**, THZ1 potently inhibits proliferation of Jurkat and Loucy T-ALL cell lines. Cell lines were treated with THZ1 or THZ1-R for 72 h. Experiments were performed in biological triplicates. Error bars show \pm standard deviation (s.d.). **c**, THZ1 and THZ1-R have different binding affinities for CDK7. LanthaScreen Eu Kinase Binding Assay was conducted at Life Technologies in a time-dependent manner. Dissociation constant (K_d) values are shown after 180 min incubation with compounds. Experiments were performed in biological triplicates. Error bars show \pm s.d.

¹Department of Cancer Biology, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. ²Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts 02115, USA. ³Whitehead Institute for Biomedical Research, 9 Cambridge Center, Cambridge, Massachusetts 02142, USA. ⁴Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ⁵Blais Proteomics Center, Dana-Farber Cancer Institute, Boston, Massachusetts 02115, USA. ⁶Department of Medicine Massachusetts General Hospital Cancer Center and Harvard Medical School, Charlestown, Massachusetts 02129, USA. ⁷Broad Institute of MIT and Harvard, 7 Cambridge Center, Cambridge, Massachusetts 02142, USA. ⁸Department of Medical Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁹Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. ¹⁰Terry Fox Laboratory, British Columbia Cancer Agency, Vancouver, British Columbia V5Z 1L3, Canada. ¹¹Department of Pediatric Oncology, Dana-Farber Cancer Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. ¹²Cancer Science Institute of Singapore, National University of Singapore, 117599 Singapore. ¹³Chemical Kinomics Research Center, Korea Institute of Science and Technology, 39-1, Hawolgok-dong, Seongbuk-gu, Seoul 136-791, Korea, and KU-KIST Graduate School of Converging Science and Technology, 145, Anam-ro, Seongbuk-gu, Seoul 136-713, Korea. ¹⁴Daegu-Gyeongbuk Medical Innovation Foundation, 2387 dalgubeol-daero, Suseong-gu, Daegu 706-010, Korea. ¹⁵Division of Hematology/Oncology, Children's Hospital, Boston, Massachusetts 02115 USA.

*These authors contributed equally to this work.

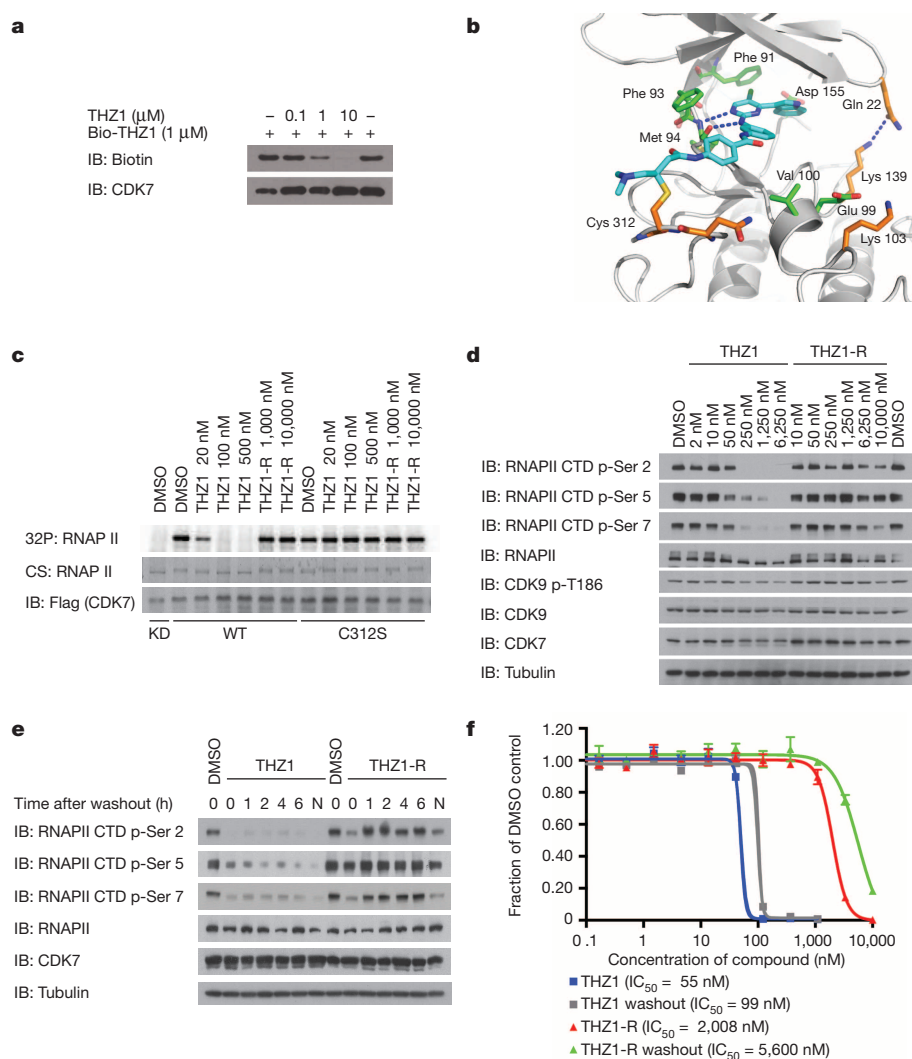


Figure 2 | THZ1 irreversibly inhibits RNAPII CTD phosphorylation by covalently targeting a unique cysteine located outside the kinase domain of CDK7. **a**, Bio-THZ1 binds irreversibly to CDK7. Recombinant CAK complex was incubated with bio-THZ1 with or without THZ1 at 37 °C for 4 h and biotinylated proteins were resolved by SDS–polyacrylamide gel electrophoresis (SDS–PAGE). IB, immunoblot. **b**, Docking model of THZ1 in the ATP-binding pocket of CDK7 (Protein Data Bank accession 1UA2). CDK7 is depicted with grey ribbons and THZ1 in turquoise. Key residues are indicated. C312 has been modelled into the crystal structure. **c**, Mutation of C312 to serine (C312S) rescues wild-type kinase activity in the presence of THZ1. HCT116 cells stably expressing Flag-tagged CDK7 proteins were treated with THZ1 or THZ1-R for 4 h. Exogenous CDK7 proteins were immunoprecipitated with Flag antibody and subjected to *in vitro* kinase assays. CS, Coomassie stain; KD, kinase-dead; WT, wild type. **d**, THZ1 inhibits RNAPII CTD phosphorylation. Jurkat cells were treated with THZ1 or THZ1-R for 4 h and proteins of interest were resolved by SDS–PAGE. **e**, THZ1, but not THZ1-R, shows irreversible inactivation of CDK7. Jurkat cells were treated with THZ1 or THZ1-R for 4 h followed by washout of inhibitor-containing medium. Cells were then allowed to grow in medium without inhibitor for 0–6 h. 'N' indicates no washout, meaning that cells were treated with compound for the duration of the experiment (10 h). **f**, Antiproliferative effects of THZ1 are impervious to inhibitor washout. Jurkat cells were treated with THZ1 or THZ1-R in dose–response format for 72 h. Experiments were performed in biological triplicates. Error bars show \pm s.d.

version of THZ1 (bio-THZ1; Fig. 1a) and demonstrated that it indeed covalently modifies CDK7 (Fig. 2a and Extended Data Fig. 1d–g). Mass spectrometry identified the site of covalent modification as C312, a residue located outside the kinase domain (Extended Data Fig. 2a–d). Inspection of the crystal structure revealed that a carboxy-terminal extension of CDK7 bearing C312 traverses the ATP cleft in the kinase domain and would be predicted to position C312 directly adjacent to the reactive acrylamide moiety of THZ1 (Fig. 2b). Mutation to serine (C312S), a less nucleophilic amino acid, prevented THZ1 from covalently binding to CDK7 and from inhibiting CDK7 activity in an irreversible fashion (Fig. 2c and Extended Data Fig. 2e). Sequence alignment of the 20-member CDK family suggests that C312 is unique to CDK7; however CDK12 and CDK13 also possess accessible cysteines within four amino acids of C312 (Extended Data Fig. 3a). Indeed, we found that THZ1 can inhibit CDK12 kinase activity at slightly higher concentrations (Extended Data Fig. 3b–f). To our knowledge, THZ1 is the first inhibitor that has been demonstrated to target a cysteine located outside of the kinase domain, which provides an unanticipated means of achieving covalent selectivity.

CDK7 kinase activity has been implicated in the regulation of both transcription, where it phosphorylates the C-terminal domain (CTD) of RNAP polymerase II (RNAPII)^{6–8} and CDK9 (ref. 9), and the cell cycle, where it functions as the CDK-activating kinase (CAK) for CDK1, 2, 4 and 6 (refs 10–14). THZ1, but not THZ1-R, completely inhibits the phosphorylation of the established intracellular CDK7 substrate RNAPII CTD at Ser 5 and Ser 7 (refs 6, 8), with concurrent loss of Ser 2 phosphorylation at 250 nM in Jurkat cells (Fig. 2d). Cellular washout experiments demonstrate

that THZ1 indeed acts in an irreversible fashion (Fig. 2e, f and Extended Data Fig. 4a–e). We observed a loss of CAK activity, as evidenced by decreased phosphorylation of the activation loops of CDK1, 2 and 9, indicating disruption of both recognized CDK7 signalling pathways in Jurkat cells (Fig. 2d and Extended Data Fig. 4f, g) and Loucy cell lines (Extended Data Fig. 4). Ectopic expression of doxycycline-inducible Flag-tagged CDK7 C312S, but not Flag–CDK7 wild type, in HeLa S3 cells restored RNAPII CTD phosphorylation (p)-Ser 5/7 to near wild-type levels at concentrations of THZ1 up to 2.5 μM, establishing C312 as a critical determinant of the cellular pharmacology of the inhibitor (Extended Data Fig. 5a, b). Additionally, Flag–CDK7 C312S expression restored CDK1/2 T-loop phosphorylation, reduced early induction of cleaved poly-ADP ribose polymerase (PARP) and restored the expression of a subset of genes, including the highly expressed transcription factors *MYC*, *KLF4*, *ID1* and *GATA2* (Extended Data Fig. 5c–e). The partial rescues of the hyperphosphorylated form of RNAPII (RNAPII⁰) and RNAPII p-Ser 2 CTD phosphorylation combined with the incomplete restoration of gene expression may result, in part, from lower-affinity cross-reactivity of THZ1 with CDK12 and 13, which are bona fide Ser 2 kinases¹⁵.

Our evidence that CDK7 inhibition leads to a reduction in RNAPII CTD phosphorylation status seems to be in conflict with evidence that inhibition of CDK7 alone is insufficient to reduce RNAPII CTD phosphorylation in HCT116 cells⁹. It is possible that covalent inhibition and reversible inhibition can engender different effects on kinase structure; we did not find evidence that THZ1 affects TFIIH or CAK complex stability (Extended Data Fig. 4h). It is also possible that inhibition of CDK12/13 (or

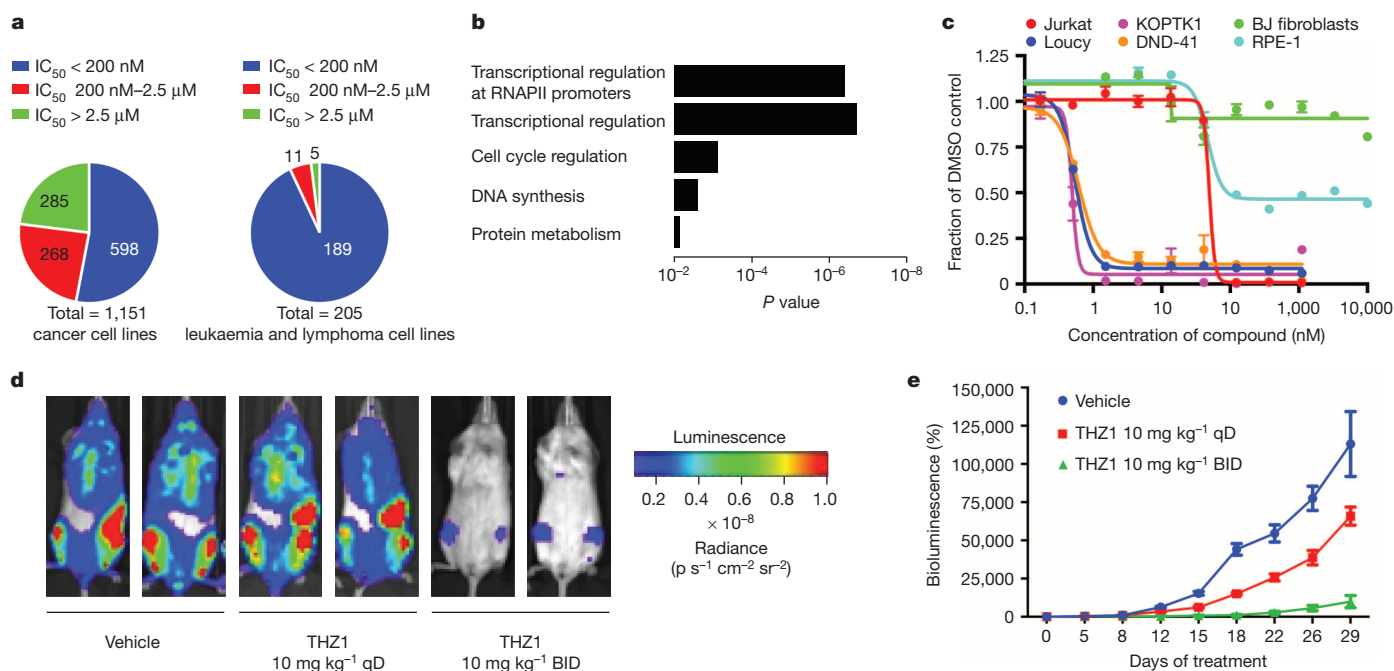


Figure 3 | THZ1 strongly reduces the proliferation and cell viability of T-ALL cell lines. **a**, THZ1 exhibits strong antiproliferative effects across a broad range of cancer cell lines from various cancer types. Cells were treated with THZ1 or DMSO vehicle for 72 h and assessed for antiproliferative effects using resazurin. **b**, Overexpression of transcriptional regulators, including (proto-) oncogenic transcription factors, is a strong predictor of cell-line sensitivity to THZ1. GO terms associated with overexpressed factors found in THZ1-sensitive cell lines. **c**, THZ1 shows a strong antiproliferative effect against T-ALL cell lines. BJ fibroblasts and RPE-1 cells are shown as normal cell lines. Cells were treated with THZ1 or DMSO vehicle for 72 h. Experiments were performed in biological triplicates. Error bars show \pm s.d. **d**, THZ1

reduces the proliferation of KOPK1 T-ALL cells in a human xenograft mouse model. Bioluminescent images of two representative mice treated with either vehicle control, 10 mg kg⁻¹ THZ1 once daily (qD), or 10 mg kg⁻¹ THZ1 twice daily (BID) for 29 days. **e**, Relative bioluminescence of mice treated with vehicle, 10 mg kg⁻¹ THZ1 once daily, or 10 mg kg⁻¹ THZ1 twice daily during the 29 days of treatment. $n = 10$ for all groups. Bioluminescence is shown relative to day 0 and is plotted as average \pm standard error of the mean. Analysis of the bioluminescence data by repeated measures two-way analysis of variance (ANOVA) reveals that the antiproliferative effect of treatment with THZ1 twice daily is highly statistically significantly different ($P < 0.0001$) as compared with the other treatments.

another undetected kinase) contributes to reduced RNAPII CTD phosphorylation, although our evidence that RNAPII CTD phosphorylation levels are restored after expression of CDK7 C312S suggests otherwise.

To understand better the breadth of antiproliferative activity of THZ1, we screened it against a diverse panel of over 1,000 cancer cell lines¹⁶. THZ1 showed broad-based activity with half-maximum inhibitory concentration (IC_{50}) values less than 200 nM against 53% of the cell lines tested (Fig. 3a and Supplementary Table 4). Elastic net regression analysis incorporating gene expression, copy number and sequence variation genomics data¹⁶ across 527 of the cell lines tested were used to identify genomic features common to sensitive cell lines. Gene ontology (GO) term enrichment analysis¹⁷ indicated a strong enrichment of (proto-) oncogenic transcription factors commonly overexpressed in cancer and factors involved in RNAPII-driven transcriptional regulation, suggesting that the dominant activity of THZ1 was through modulation of transcription (Fig. 3b and Supplementary Table 5).

In agreement with the net elastic regression analysis, T-ALL cell lines that show characteristic misregulation of T-cell lineage-specific transcription factors were broadly sensitive to THZ1, but not to THZ1-R (Fig. 3c, Extended Data Fig. 6a and Supplementary Table 4). Treatment of T-ALL cell lines with THZ1 caused decreased cellular proliferation and an increase in apoptotic index with concomitant reduction in anti-apoptotic proteins, most notably MCL1 and XIAP (Extended Data Figs 6 and 7). These strong antiproliferative responses induced at sub-effective doses of THZ1 suggest that T-ALL cells may be particularly sensitive to small perturbations in transcription and CDK7 kinase function. Indeed, THZ1 potently reduced the viability of patient-derived T-ALL and chronic lymphocytic leukaemia (CLL) cells (Extended Data Fig. 8a, b). Moreover, THZ1 exhibited efficacy in a bioluminescent xenografted mouse model

using the human T-ALL cell-line KOPK1, when dosed twice daily at 10 mg kg⁻¹ (Fig. 3d, e, Extended Data Fig. 8 and Supplementary Table 6). Importantly, THZ1 was well tolerated at these doses with no observable body weight loss or behavioural changes (Extended Data Fig. 8f), suggesting that it caused no overt toxicity in the animals. These results were mirrored in cell culture with non-transformed BJ fibroblast and retinal pigment epithelial (RPE-1) cells responding to relatively high doses of THZ1 by undergoing cell-cycle arrest rather than initiating apoptosis or cell death, further suggesting that normal cells might tolerate transcriptional disruption (Extended Data Fig. 9).

CDK7 is a component of the general transcription factor IIH (TFIIH) complex^{18–20}, so we next investigated how THZ1 treatment affects genome-wide gene expression. We chose Jurkat T-ALL cells for these studies because it is a well-studied T-ALL cell-line model with a defined core transcriptional regulatory circuitry consisting of key transcription factors, which is also found in human T-ALL primagrafts²¹. Treatment with 250 nM THZ1, but not THZ1-R, led to progressive reduction in global steady-state messenger RNA levels over time, with 75% and 96% of mRNAs showing greater than two-fold reduction by 6 and 12 h, respectively (Fig. 4a, Extended Data Fig. 10a and Supplementary Table 7). Consistent with global downregulation of mRNA transcripts, 250 nM THZ1 reduced RNAPII occupancy genome wide at both promoters and gene bodies (Fig. 4b). By comparison, Flavopiridol reduced RNAPII density only across gene bodies (Fig. 4b). This is consistent with the model that CDK7 regulates RNAPII initiation and pausing whereas CDK9 regulates pause release leading to processive elongation^{8,9,18–20,22–24}.

Although 250 nM THZ1 inhibits global transcription, we found that some cancer cell lines, particularly T-ALL, are sensitive to considerably lower concentrations of THZ1. We postulated that the expression of certain genes might be especially sensitive to low doses of THZ1 and

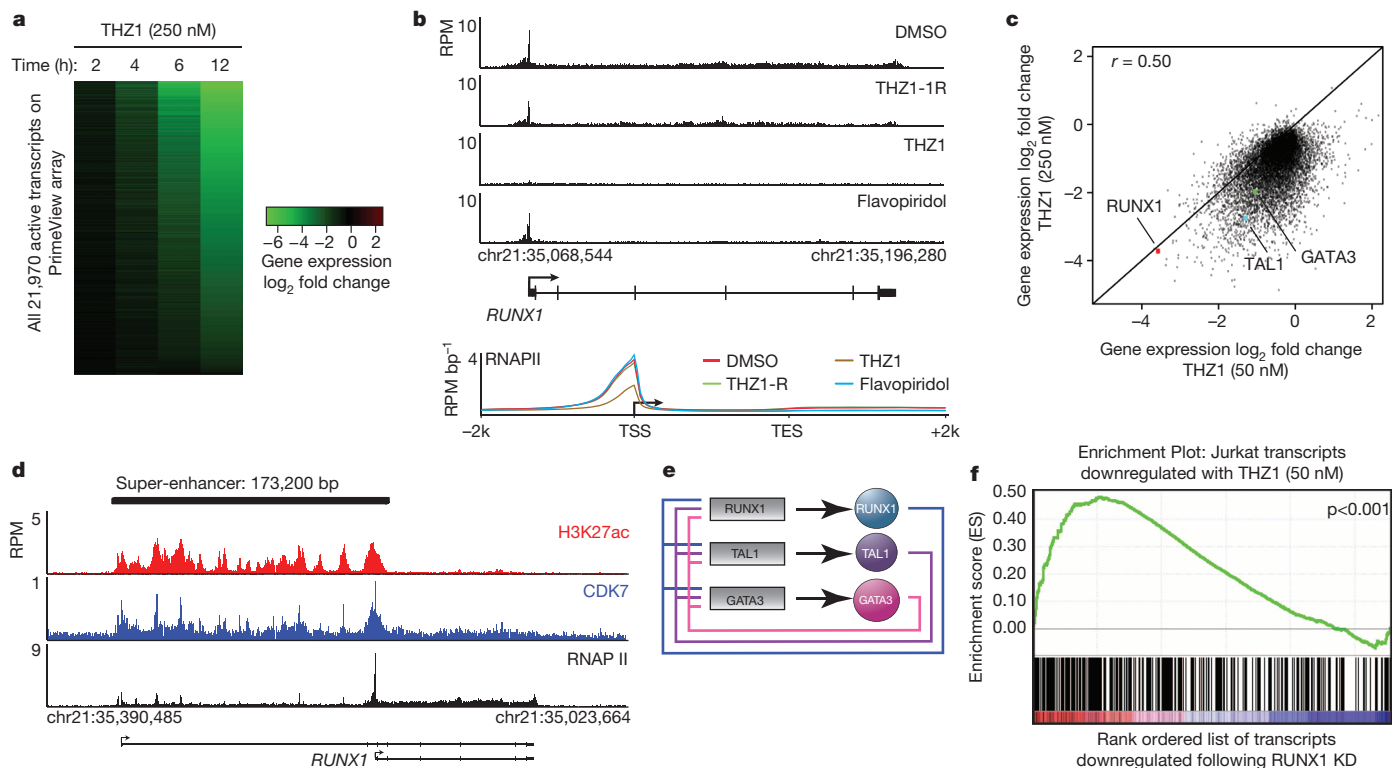


Figure 4 | THZ1 preferentially downregulates Jurkat core transcriptional circuitry. **a**, THZ1 treatment globally downregulates steady-state mRNA levels in a time-dependent manner. Jurkat cells were treated with THZ1 (250 nM) for the indicated amounts of time. Heatmaps display the log₂ fold change in gene expression versus DMSO for the 21,970 transcripts expressed at 12 h in DMSO. **b**, THZ1 reduces RNAPII occupancy across promoters and gene bodies. Gene tracks of RNAPII ChIP-seq occupancy at *RUNX1* after the indicated treatments (top). Metagene representation of global RNAPII occupancy at promoters and gene bodies (bottom). Average background subtracted chromatin immunoprecipitation followed by sequencing (ChIP-seq) signal in 22,310 genes expressed in 6 h DMSO conditions in units of reads per million per base pair (RPM bp⁻¹). Signal of ChIP-seq occupancy is in units of RPM per bin. All treatments were 6 h with 250 nM of THZ1, THZ1-R or Flavopiridol. The metagene representation spans 2 kb upstream from the transcription start site (TSS) to 2 kb downstream from the transcription end

site (TES). **c**, THZ1 treatment delineates a subset of transcripts equally sensitive to low (50 nM) and high (250 nM) dose THZ1. Log₂ fold change in gene expression for 50 nM (x-axis) and 250 nM THZ1 (y-axis) after a 4 h treatment. Pearson coefficient $r = 0.50$. **d**, Gene tracks of H3K27ac (top), CDK7 (middle) and RNAPII (bottom) ChIP-seq occupancy at the TSS, gene body, and a previously described enhancer region in the first intron of *RUNX1* (ref. 29). Total ChIP-seq signal is in units of RPM per bin. **e**, Positive interconnected autoregulatory loop formed by *RUNX1*, *TAL1* and *GATA3*. Genes are represented by rectangles, and proteins are represented by ovals²¹. **f**, Transcripts downregulated by low-dose THZ1 are enriched for transcripts downregulated after *RUNX1* knockdown. Gene set enrichment analysis of top 500 transcripts downregulated after a 4 h treatment with THZ1 (50 nM) in comparison to transcripts following a *RUNX1* knockdown²¹. Gene set enrichment analysis-supplied P value < 0.001.

therefore have a key role in driving the cellular response. Indeed, we found that transcripts for only a subset of genes were substantially affected by treatment with 50 nM THZ1, with that for *RUNX1* among the most profoundly affected (Fig. 4c). There are at least two reasons that low-dose THZ1 treatment might cause a preferential loss of *RUNX1* expression. Tumour-cell oncogenes can acquire super-enhancers, which drive high-level expression yet can be especially sensitive to perturbation^{25–28}. Super-enhancer analysis in Jurkat cells revealed that *RUNX1* contains an exceptionally large super-enhancer domain containing a previously described haematopoietic-cell-specific enhancer (Fig. 4d, Extended Data Fig. 10b–d and Supplementary Table 8)²⁹. In addition, *RUNX1* forms a core regulatory circuitry with two additional transcription factors that have prominent roles in leukaemia biology, *TAL1* and *GATA3* (Fig. 4e)²¹. These factors autoregulate their own gene expression while simultaneously regulating many other genes that comprise the active gene expression program of Jurkat cells. Treatment with 50 nM THZ1 led to a significant reduction in both the transcript and protein levels of *RUNX1*, *TAL1* and *GATA3* (Extended Data Fig. 10e, f). Loss of the *RUNX1*-driven transcriptional program is probably key to the response to low-dose THZ1 treatment, as gene set enrichment analysis revealed that the Jurkat transcripts downregulated by 50 nM THZ1 were enriched in transcripts similarly downregulated after *RUNX1* depletion using short hairpin (sh)RNA (Fig. 4f).

We have reported the discovery and characterization of a covalent inhibitor of CDK7, THZ1. THZ1 uses a unique mechanism, combining ATP-site and allosteric covalent binding, as a means of attaining potency and selectivity for CDK7. This mechanistic insight should be useful for designing next-generation inhibitors of CDKs, for which high sequence and shape homology in the ATP pocket has posed a formidable challenge to achieving selectivity with conventional ATP-competitive inhibitors. THZ1 showed potent antiproliferative activity on T-ALL cell lines and other blood cancers, in which oncogenic transcription factors feature prominently in the disease state. In Jurkat cells, low-dose THZ1 had a profound effect on a small subset of genes, including the key regulator *RUNX1*, thus contributing to subsequent loss of the greater gene expression program and cell death. Identification of additional cancer cell lines whose gene expression programs show vulnerability to THZ1 or other transcriptional inhibitors should delineate other cancers that are susceptible to perturbation of transcription.

METHODS SUMMARY

T-ALL culture conditions. Jurkat, Loucy, KOPTK1 and DND-41 cell lines were grown in RPMI-1640 supplemented with 10% fetal bovine serum (FBS) and 1% glutamine. All cell lines were cultured at 37 °C in a humidified chamber in the presence of 5% CO₂, unless otherwise noted.

Inhibitor treatment experiments. Time-course experiments such as those described in Extended Data Fig. 5a were conducted to determine the minimal time required for full inactivation of CDK7. Cells were treated with THZ1, THZ1-R or dimethylsulphoxide (DMSO) for 0–6 h to assess the effect of time on the THZ1-mediated inhibition of RNAPII CTD phosphorylation. For subsequent experiments cells were treated with compounds for 4 h as determined by the time-course experiment described earlier, unless otherwise noted. For inhibitor washout experiments (Fig. 2e, f and Extended Data Fig. 5) cells were treated with THZ1, THZ1-R or DMSO for 4 h. Medium containing inhibitors was subsequently removed to effectively ‘washout’ the compound and the cells were allowed to grow in the absence of inhibitor. For each experiment, lysates were probed for RNAPII CTD phosphorylation and other specified proteins.

High-throughput cell-line panel viability assay. Cells were seeded in 384-well microplates at ~15% confluency in medium with 5% FBS and penicillin/streptavidin. Cells were treated with THZ1 or DMSO for 72 h and cell viability was determined using resazurin.

RNA extraction and synthetic RNA spike-in. Total RNA and sample preparation was performed as previously described³⁰. Briefly, after inhibitor treatment cell number was determined, total RNA was isolated, and ERCC RNA Spike-In Mix (Ambion, catalogue no. 4456740) was added to total RNA relative to cell number.

Expanded protocols and synthetic chemistry schemes can be found in Supplementary Information.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 September 2013; accepted 14 April 2014.

Published online 22 June 2014.

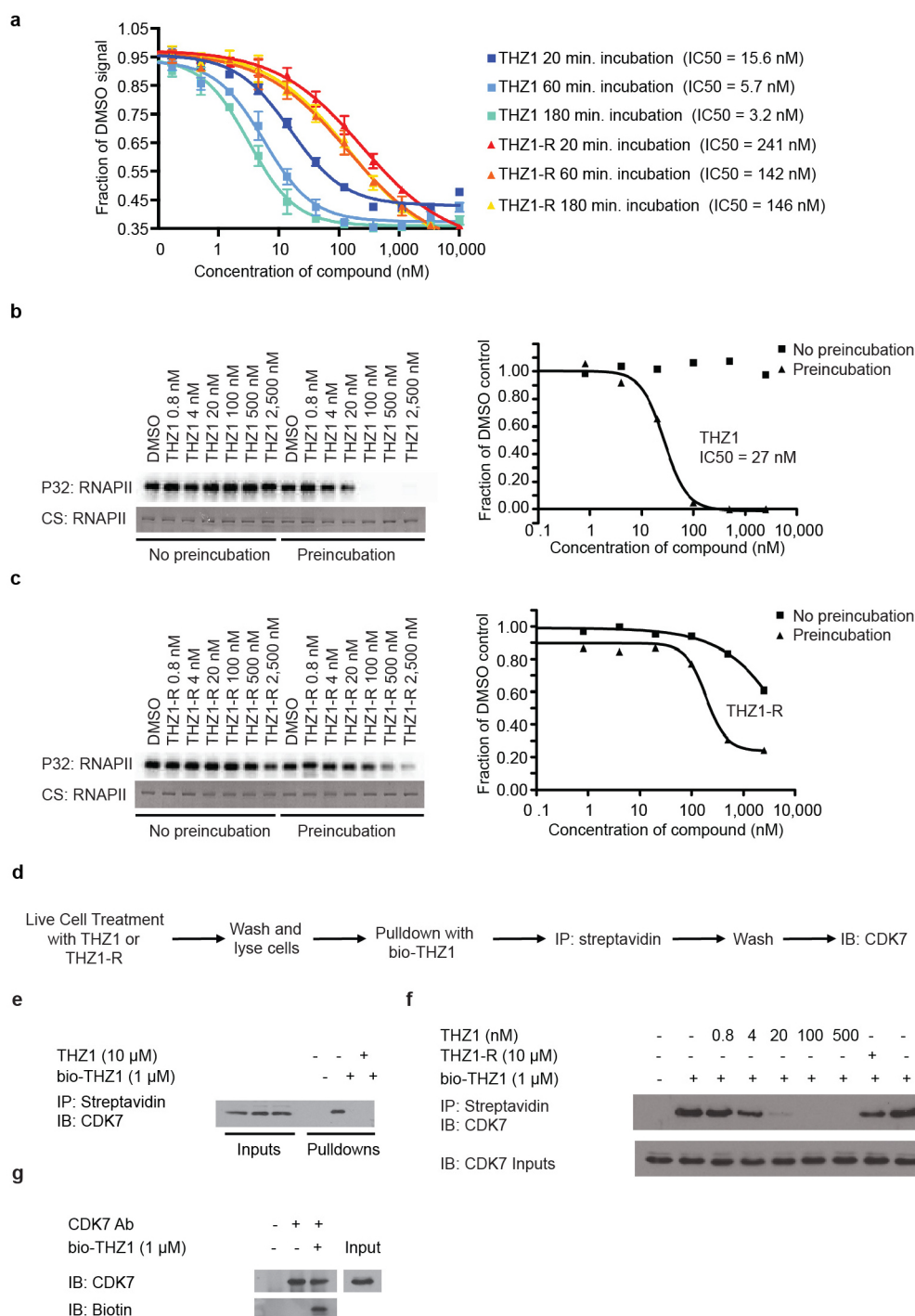
- O’Neil, J. & Look, A. T. Mechanisms of transcription factor deregulation in lymphoid cell transformation. *Oncogene* **26**, 6838–6849 (2007).
- Berg, T. Inhibition of transcription factors with small organic molecules. *Curr. Opin. Chem. Biol.* **12**, 464–471 (2008).
- Lu, Q. *et al.* Perspectives on the discovery of small-molecule modulators for epigenetic processes. *J. Biomol. Screen.* **17**, 555–571 (2012).
- Sansó, M. & Fisher, R. P. Pause, play, repeat: CDKs push RNAP II’s buttons. *Transcription* **4**, 146–152 (2013).
- Patricelli, M. P. *et al.* Functional interrogation of the kinome using nucleotide acyl phosphates. *Biochemistry* **46**, 350–358 (2007).
- Akhtar, M. S. *et al.* TFIIF kinase places bivalent marks on the carboxy-terminal domain of RNA polymerase II. *Mol. Cell* **34**, 387–393 (2009).
- Drapkin, R., Le Roy, G., Cho, H., Akoulitchev, S. & Reinberg, D. Human cyclin-dependent kinase-activating kinase exists in three distinct complexes. *Proc. Natl Acad. Sci. USA* **93**, 6488–6493 (1996).
- Glover-Cutter, K. *et al.* TFIIF-associated Cdk7 kinase functions in phosphorylation of C-terminal domain Ser7 residues, promoter-proximal pausing, and termination by RNA polymerase II. *Mol. Cell Biol.* **29**, 5455–5464 (2009).
- Larochelle, S. *et al.* Cyclin-dependent kinase control of the initiation-to-elongation switch of RNA polymerase II. *Nature Struct. Mol. Biol.* **19**, 1108–1115 (2012).
- Fisher, R. P. & Morgan, D. O. A novel cyclin associates with MO15/CDK7 to form the CDK-activating kinase. *Cell* **78**, 713–724 (1994).
- Larochelle, S. *et al.* Requirements for Cdk7 in the assembly of Cdk1/cyclin B and activation of Cdk2 revealed by chemical genetics in human cells. *Mol. Cell* **25**, 839–850 (2007).
- Mäkelä, T. P. *et al.* A cyclin associated with the CDK-activating kinase MO15. *Nature* **371**, 254–257 (1994).
- Schachter, M. M. *et al.* A Cdk7-Cdk4 T-loop phosphorylation cascade promotes G1 progression. *Mol. Cell* **50**, 250–260 (2013).
- Solomon, M. J., Lee, T. & Kirschner, M. W. Role of phosphorylation in p34cdc2 activation: identification of an activating kinase. *Mol. Biol. Cell* **3**, 13–27 (1992).
- Bartkowiak, B. *et al.* CDK12 is a transcription elongation-associated CTD kinase, the metazoan ortholog of yeast Ctk1. *Genes Dev.* **24**, 2303–2316 (2010).
- Garnett, M. J. *et al.* Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* **483**, 570–575 (2012).
- Huang da, W., Sherman, B. T. & Lempicki, R. A. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* **37**, 1–13 (2009).
- Feaver, W. J., Svejstrup, J. Q., Henry, N. L. & Kornberg, R. D. Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIF/TFIIK. *Cell* **79**, 1103–1109 (1994).
- Serizawa, H. *et al.* Association of Cdk-activating kinase subunits with transcription factor TFIIF. *Nature* **374**, 280–282 (1995).
- Shiekhattar, R. *et al.* Cdk-activating kinase complex is a component of human transcription factor TFIIF. *Nature* **374**, 283–287 (1995).
- Sanda, T. *et al.* Core transcriptional regulatory circuit controlled by the TAL1 complex in human T cell acute lymphoblastic leukemia. *Cancer Cell* **22**, 209–221 (2012).
- Rahl, P. B. *et al.* c-Myc regulates transcriptional pause release. *Cell* **141**, 432–445 (2010).
- Watanabe, Y. *et al.* Modulation of TFIIF-associated kinase activity by complex formation and its relationship with CTD phosphorylation of RNA polymerase II. *Genes Cells* **5**, 407–423 (2000).
- Yamada, T. *et al.* P-TEFb-mediated phosphorylation of hSpt5 C-terminal repeats is critical for processive transcription elongation. *Mol. Cell* **21**, 227–237 (2006).
- Chapuy, B. *et al.* Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell* **24**, 777–790 (2013).
- Hnisz, D. *et al.* Super-enhancers in the control of cell identity and disease. *Cell* **155**, 934–947 (2013).
- Lovén, J. *et al.* Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **153**, 320–334 (2013).
- Shi, J. *et al.* Role of SWI/SNF in acute leukemia maintenance and enhancer-mediated Myc regulation. *Genes Dev.* **27**, 2648–2662 (2013).
- Nottingham, W. T. *et al.* Runx1-mediated hematopoietic stem-cell emergence is controlled by a Gata/Ets/SCL-regulated enhancer. *Blood* **110**, 4188–4197 (2007).
- Lovén, J. *et al.* Revisiting global gene expression analysis. *Cell* **151**, 476–482 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank members of the Gray and Young laboratories for helpful discussions; D. Orlando, L. Lawton and L. Anders for advice; and C. Thoreen and D. Sabatini, as well as S. Cheng and G. Morin, for reagents. We thank K. Jones and N. Kohl for performing mouse studies and K. Jones and C. Christensen for prepping mouse tissues. We thank S. Riddle for performing LanthaScreen Eu Kinase Assays. This work was supported by the National Institutes of Health (R01 CA130876-04 and U54 HG006097-02 to N.S.G.; CA178860-01 and P01 NS047572-10 to J.A.M.; HG002668 and CA109901 to R.A.Y.), and the American Cancer Society Postdoctoral Fellowship 120272-PF-11-042-01-DMC (P.B.R.). J.R.B. is a Scholar in Clinical Research of the Leukemia Lymphoma Society and is supported by the Translational Research Program of the Leukemia Lymphoma Society and by the American Cancer Society.

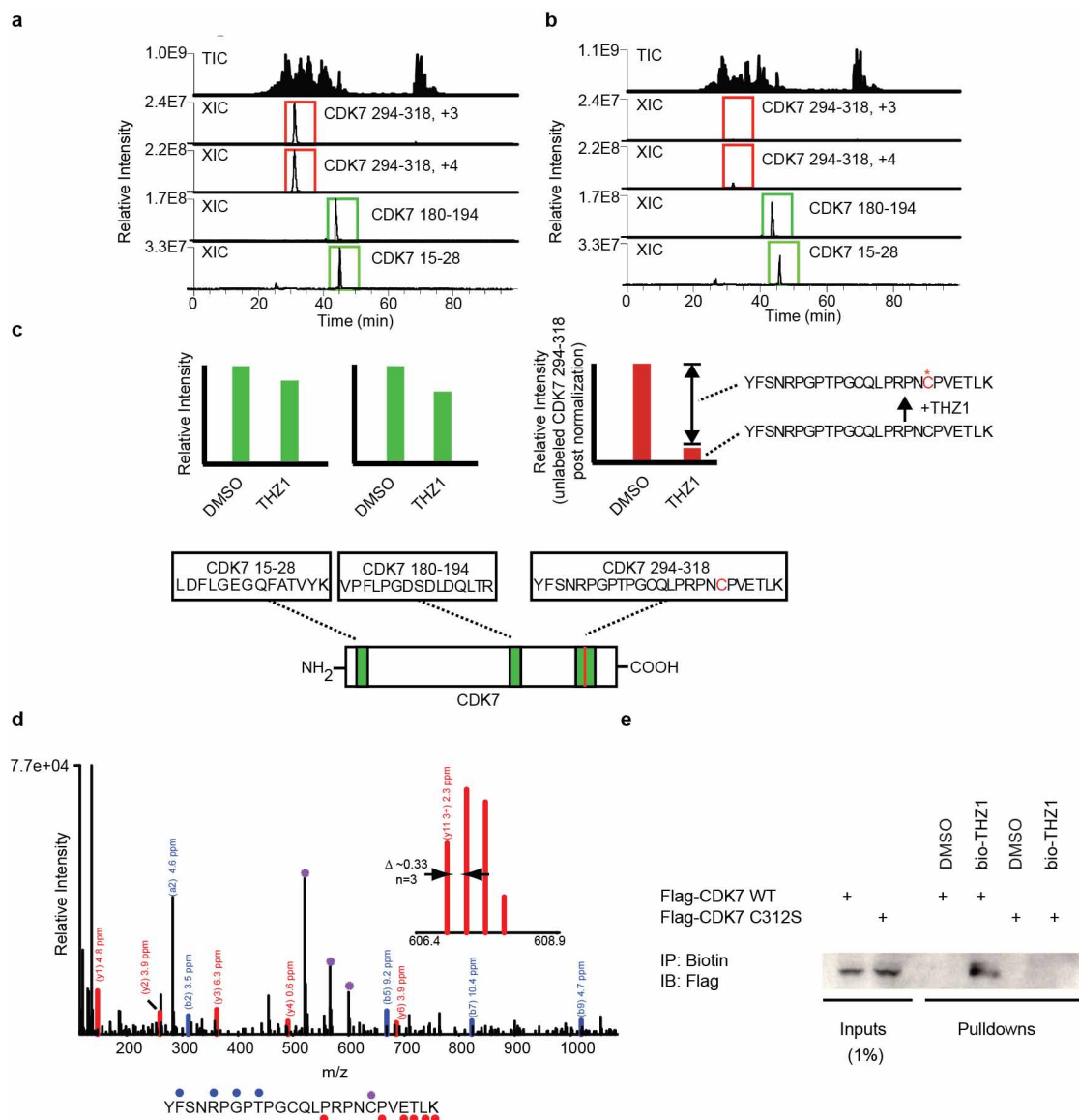
Author Contributions N.S.G., N.K. and T.Z. conceived the project. N.S.G. and T.Z. conceived and directed the chemical synthesis of THZ1 and its analogues with input from T.Si. T.Z. performed chemical synthesis and small-molecule structure determination. R.A.Y., N.S.G., P.B.R. and N.K. conceived genomics studies. N.K., P.B.R., J.R.B., C.H.B., N.S.G. and R.A.Y. designed biological experimental research with input from J.R., B.J.A., D.M., T.Sa., T.L., A.P.W. and C.S.M. N.K., P.B.R., J.R., A.D., B.T., C.E.J. and N.M.H. performed experimental biological research. S.B.F. designed and performed protein mass spectrometry on THZ1/CDK7 adducts with input from J.A.M. S.R. and A.A. performed elastic net regression analysis. N.D.K. performed molecular modelling studies with input from T.Sa. B.J.A. designed and performed genomics data analyses. N.K., N.S.G. and R.A.Y. co-wrote the paper. All authors edited the manuscript.

Author Information Sequencing and expression data have been deposited in the Gene Expression Omnibus under accession number GSE50625. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to N.S.G. (nathanael_gray@dfci.harvard.edu) or R.A.Y. (young@wi.mit.edu).



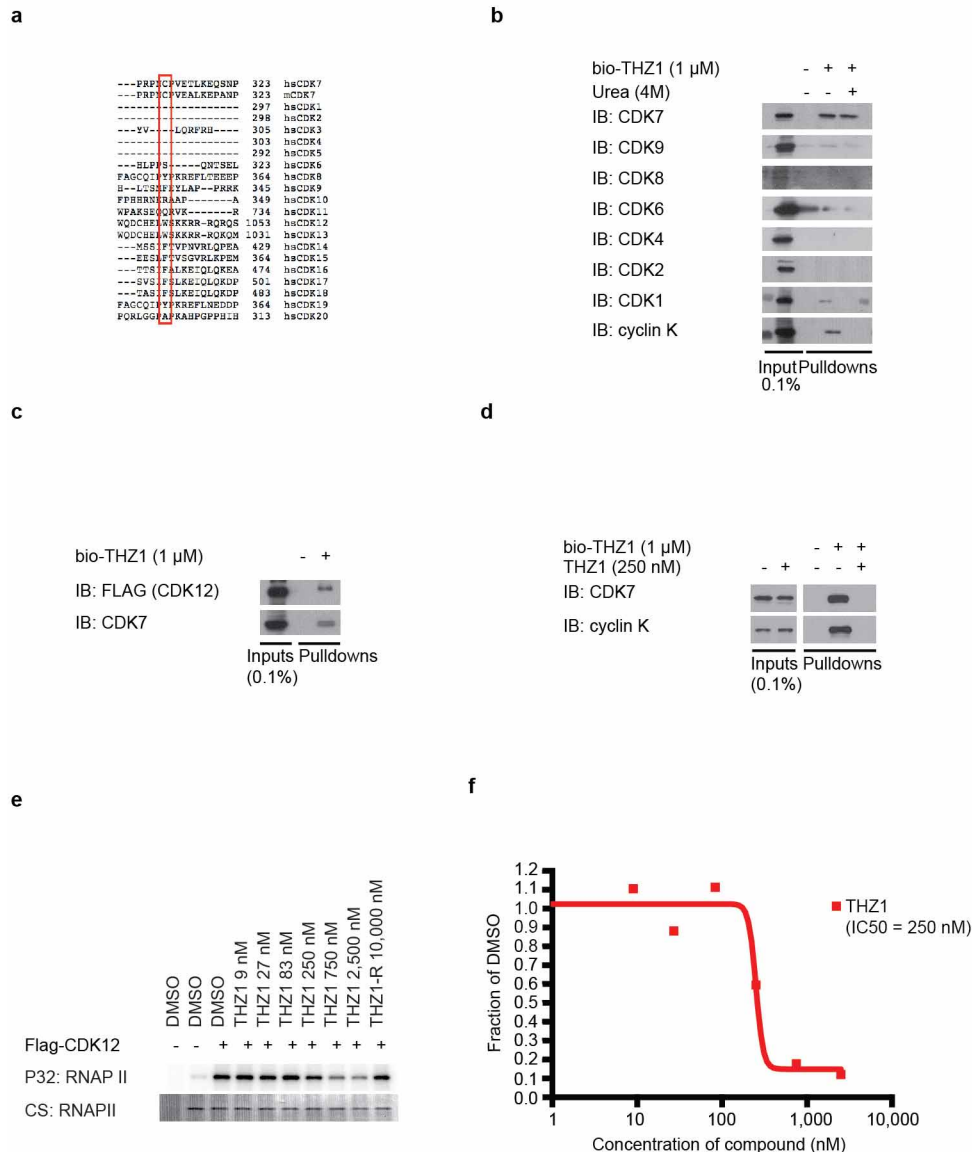
Extended Data Figure 1 | THZ1 demonstrates time-dependent inhibition of CDK7 *in vitro* and covalent binding of intracellular CDK7. **a**, THZ1 but not THZ1-R shows time-dependent inhibition. Lanthascreen Eu Kinase Binding Assay was conducted at Life Technologies in a time-dependent manner (20, 60 and 180 min), demonstrating that THZ1 but not THZ1-R shows time-dependent inhibition of CDK7. **b**, **c**, Pre-incubation of THZ1 increases CDK7 inhibitory activity *in vitro*. Recombinant CAK complex was incubated with THZ1 (**b**) or THZ1-R (**c**) in a dose-response format with or without pre-incubation before ATP (25 μM) addition. The kinase reaction was then allowed to proceed for 45 min at 30 °C. **d**, Workflow of bio-THZ1 pull-down competition experiment. **e**, Bio-THZ1 pulls down CDK7 from cellular

lysates. Loucy cellular lysates were incubated with bio-THZ1 (1 μM) with or without THZ1 (10 μM) and streptavidin-precipitated proteins were probed for CDK7. IB, immunoblot. **f**, Free intracellular THZ1 competes in a dose-dependent manner for bio-THZ1 binding to CDK7. Loucy cells were treated with increasing concentrations of THZ1 or with 10 μM THZ1-R for 4 h. Cellular lysates were incubated with bio-THZ1 and processed as indicated in **a**, **g**. Bio-THZ1 labels CDK7 in lysates. Loucy cellular lysates were incubated with bio-THZ1 at 4 °C for 12 h followed by immunoprecipitation of CDK7 at 4 °C for 3 h. Precipitated proteins were washed and probed with horseradish peroxidase (HRP)-conjugated streptavidin.



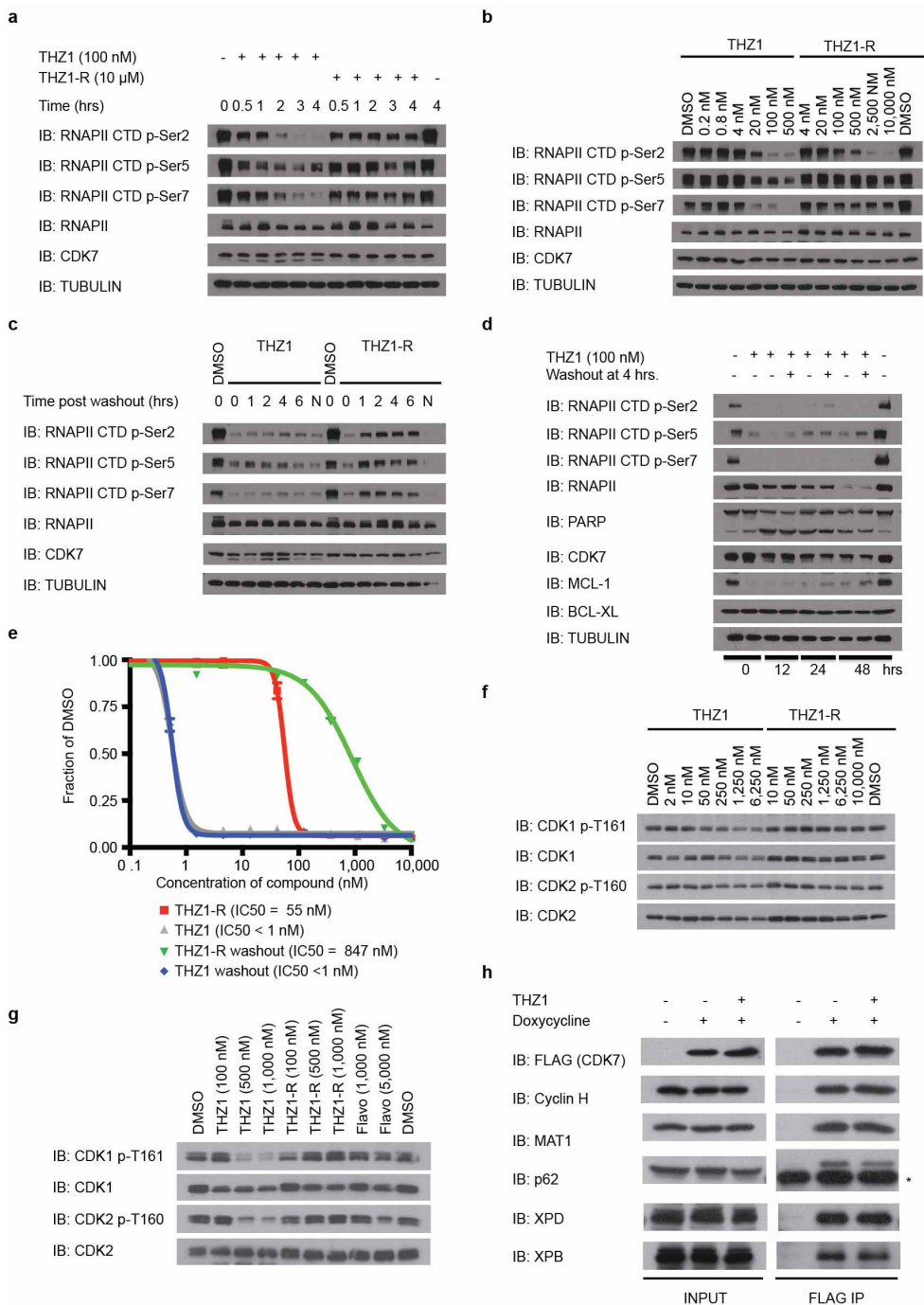
Extended Data Figure 2 | THZ1 covalently binds CDK7 C312. **a, b**, Total ion chromatograms (TIC) and extracted ion chromatograms (XIC) for CDK7 peptides recorded during analysis of CAK complexes treated with DMSO (**a**) or THZ1 (**b**). **c**, Efficiency of labelling was estimated to be approximately 85%, as gauged by the reduction in signal of triply and quadruply charged YFSNRPGPTPGCQLPRNPVETLK ions (residues 294–318). The peptides VPFLPGDSDLQLTR (residues 180–194) and LDLFGEGQFATVYK (residues 15–28) were used for normalization. **d**, Orbitrap HCD tandem mass spectrometry (MS/MS) spectrum of a quadruply charged CDK7-derived peptide (residues 294–318) labelled by THZ1 at C312. Fragment ions containing the peptide C terminus (γ -type) or N terminus (β -type), along with

the associated mass errors are shown in red and blue, respectively. Fragment ions marked by an asterisk contain the inhibitor and have the expected heavy isotope contribution from chlorine. The site of labelling was determined to be C312 (as opposed to C305) on the basis of fragment ions observed in additional MS/MS spectra (for example, $\gamma 11^{3+}$ observed with <3 p.p.m. mass error by fragmentation of the +6 charged precursor; see inset mass spectrum). **e**, C312S mutation eliminates THZ1 covalent binding. Cellular lysates from HCT116 cells expressing either Flag-CDK7 wild type or C312S were incubated with bio-THZ1 for 12 h at 4 °C and then at room temperature for 3 h to facilitate covalent binding. Precipitated proteins were then probed for the presence of Flag-tagged CDK7.



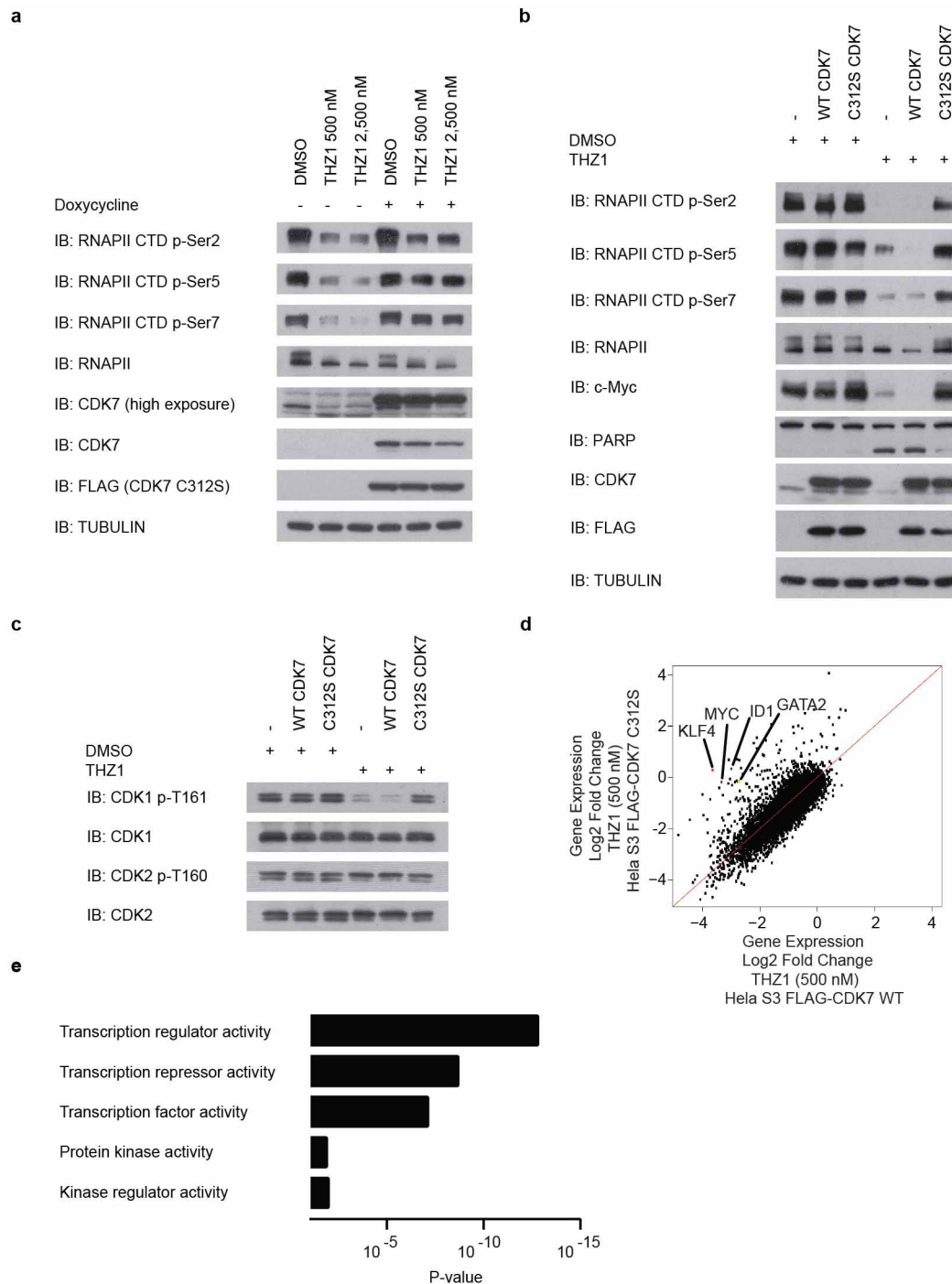
Extended Data Figure 3 | THZ1 inhibits CDK12 but at higher concentrations compared with CDK7. **a**, Protein sequence alignment of the C-terminal regions of all human (hs) CDKs and mouse (m) CDK7 using Uniprot default settings. Note that the canonical cell-cycle CDKs 1, 2 and 4, as well as 5, do not have C-terminal domains that align to the equivalent position of CDK7 C312 and therefore do not show aligned sequence in this region. **b**, Bio-THZ1 covalently pulls down CDK7 from cellular lysates. Jurkat cellular lysates were incubated with bio-THZ1 (1 μ M) at 4 $^{\circ}$ C for 12 h and for 2 h at room temperature. Precipitated proteins were washed with or without urea (4 M), here used as a denaturing agent, and probed for the indicated CDKs. **c**, Bio-THZ1 pulls down Flag-CDK12 from lysates. Lysates from 293A cells stably expressing Flag-tagged wild-type CDK12 were incubated with bio-THZ1

(1 μ M) at 4 $^{\circ}$ C for 12 h and for 2 h at room temperature. Immunoprecipitated proteins were probed with Flag antibody to recognize CDK12 or with CDK7 antibody. **d**, Bio-THZ1 pulls down cyclin K from cellular lysates. Jurkat cellular lysates were incubated with bio-THZ1 (1 μ M) at 4 $^{\circ}$ C for 12 h and for 2 h at room temperature. Precipitated proteins were probed for the indicated proteins. **e**, THZ1 inhibits CDK12 in an *in vitro* kinase assay. 293A cells stably expressing Flag-tagged wild-type CDK12 were treated with THZ1 or THZ1-R for 4 h. Exogenous CDK12 was immunoprecipitated from cellular lysates using Flag antibody. Precipitated proteins were washed and subjected to *in vitro* kinase assays at 30 $^{\circ}$ C for 30 min using the large subunit of RNAPII (RPB1) as substrate and 25 μ M ATP. CS, Coomassie stain. **f**, Quantification of *in vitro* kinase assay conducted in **d**.



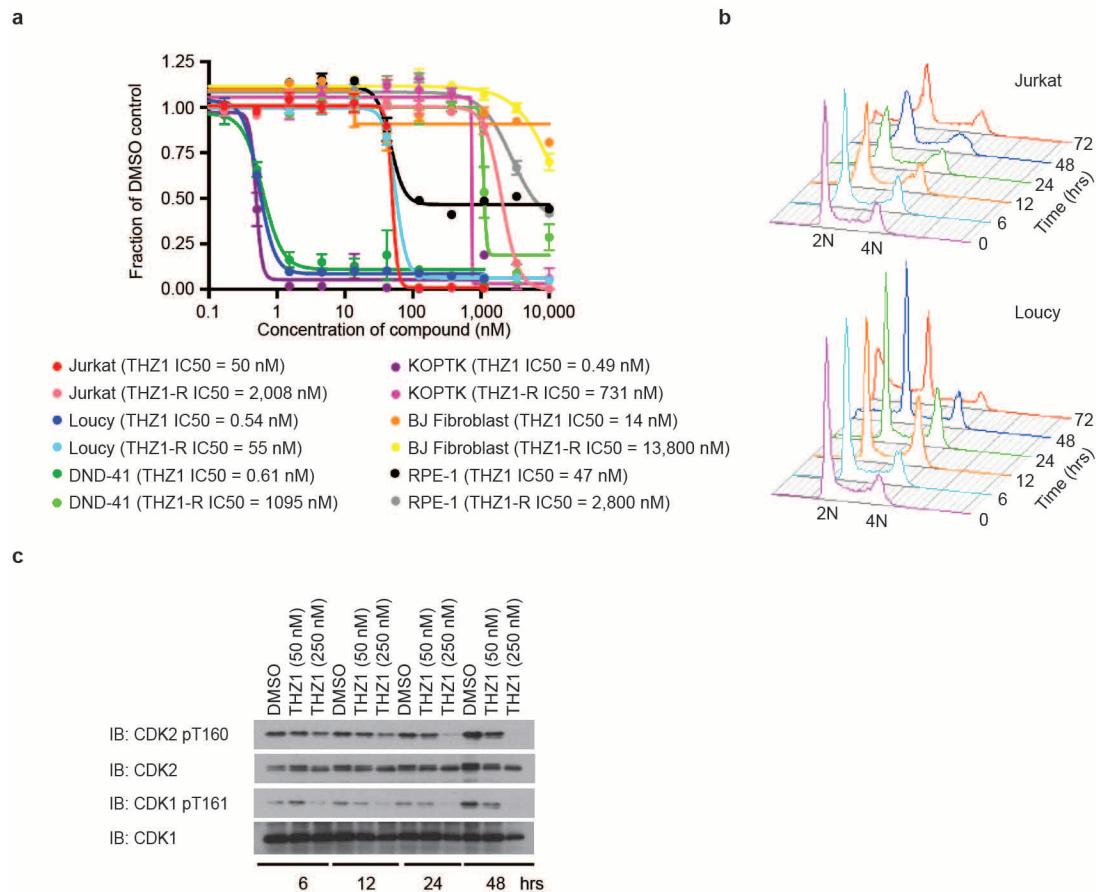
Extended Data Figure 4 | THZ1 irreversibly inhibits RNAPII CTD and CAK phosphorylation. **a**, THZ1 exhibits time-dependent inactivation of intracellular CDK7. Loucy cells were treated with THZ1 or THZ1-R for 0–4 h. At each time point, cells were harvested, lysed and the cellular lysates were probed with antibodies against the specified proteins. **b**, THZ1 inhibits RNAPII CTD phosphorylation. Loucy cells were treated with THZ1 or THZ1-R for 4 h. Cellular lysates were then probed with antibodies recognizing the Ser 2, Ser 5 and Ser 7 CTD RNAPII phospho-epitopes. **c**, Loucy cells were treated with THZ1 or THZ1-R for 4 h followed by washout of inhibitor-containing medium. Cells were allowed to grow in medium without inhibitor for 0–6 h. At each time point cells were lysed and the cellular lysates were probed with antibodies against the specified proteins. 'N' indicates cells for which medium was never washed out. **d**, Apoptotic signalling is maintained despite washout of THZ1. Loucy cells were treated with THZ1 or THZ1-R for 4 h followed by washout of inhibitor-containing medium, at which point cells were allowed to grow in medium with or without inhibitor for 0–48 h. At each time point, cells were lysed and the cellular lysates were probed with antibodies against the specified

proteins. **e**, Antiproliferative effects of THZ1 are impervious to inhibitor washout. Loucy cells were treated with THZ1 or THZ1-R in a dose-response format for 72 h. Antiproliferative effects were determined using CellTiter-Glo analysis. **f**, THZ1 reduces the T-loop phosphorylation status of CDK1 and CDK2 in Jurkat cells over a 3 h exposure. Asynchronous cells were treated with increasing concentrations of THZ1 or THZ1-R for 3 h. Cellular lysates were then probed with antibodies against the indicated proteins or phosphoproteins. **g**, THZ1, but not THZ1-R, completely inhibits T-loop phosphorylation of CDK1 and CDK2 after treatment over one cell cycle. Loucy cells were treated with THZ1, THZ1-R, Flavopiridol or DMSO vehicle at the indicated concentrations for 24 and 14 h, respectively (roughly one cell cycle). Cell lysates were harvested and probed with antibodies against the specified proteins or phosphoproteins. **h**, HeLa S3 cells stably expressing Flag-tagged wild-type CDK7 were treated with THZ1 (1 μ M) or DMSO vehicle for 5 h with or without the presence of doxycycline. Proteins were immunoprecipitated using Flag antibody. Precipitated proteins were probed using the indicated antibodies. Asterisk indicates heavy chain from IgG antibody.



Extended Data Figure 5 | Mutation of CDK7 C312 to serine rescues Ser 5/7 and partially rescues Ser 2 RNAPII CTD phosphorylation. **a**, Expression of C312S rescues Ser 5/7 and partially rescues Ser 2 RNAPII CTD phosphorylation. HeLa S3 cells stably carrying a doxycycline-inducible Flag-CDK7 C312S construct were treated with THZ1 or DMSO for 5 h with or without the presence of doxycycline. Cellular lysates were then probed for the indicated proteins. **b**, Phenotypic rescue is specific to the C312S mutation, as rescue is not achieved with overexpression of Flag-CDK7 wild type (WT). HeLa S3 cells stably carrying doxycycline-inducible Flag-CDK7 wild-type and C312S constructs (or empty vector) were treated with THZ1 or DMSO for 5 h in the presence of doxycycline. **c**, Expression of C312S largely restores

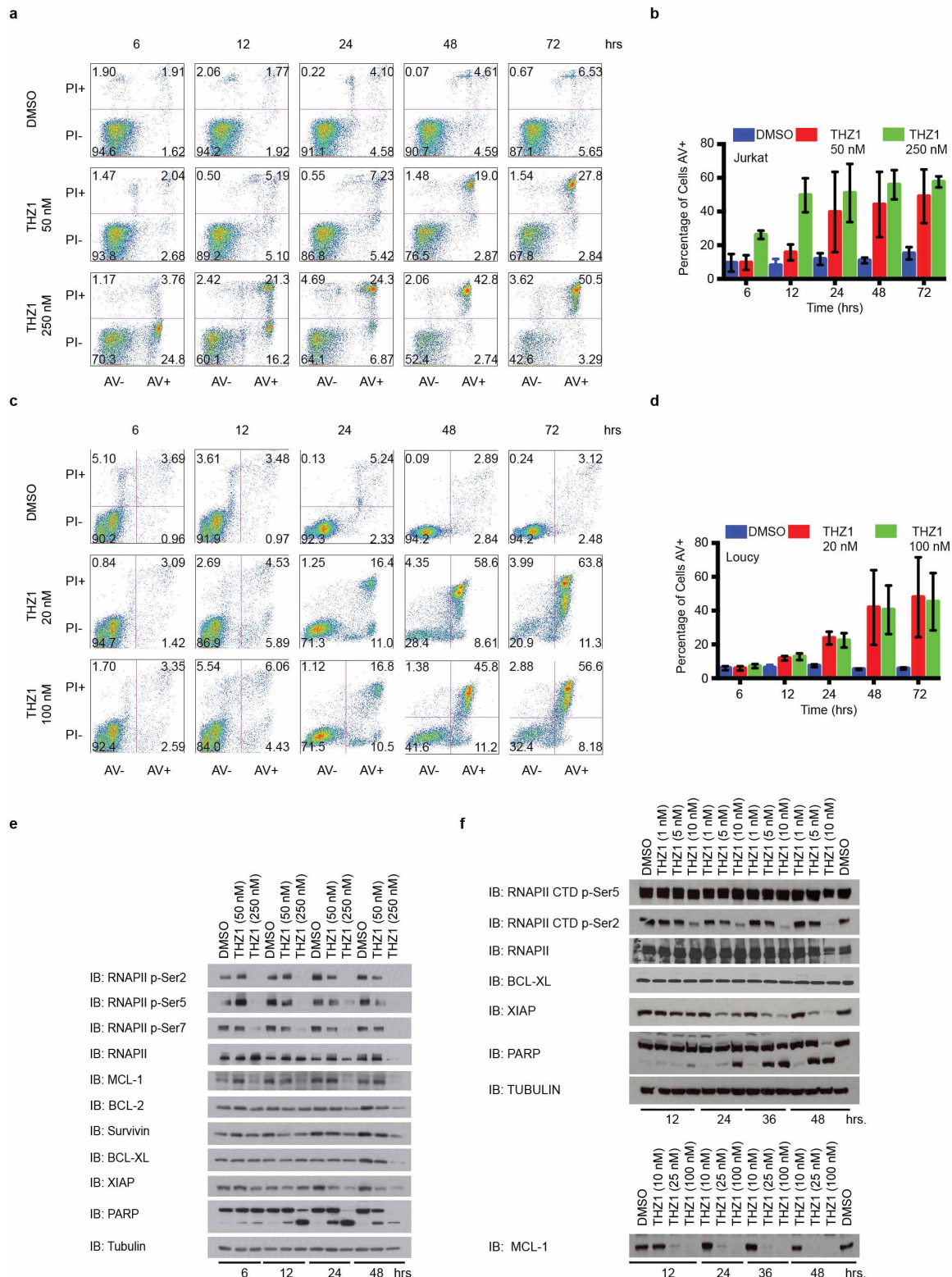
CDK1/2 T-loop phosphorylation. HeLa S3 cells stably carrying a doxycycline-inducible Flag-CDK7 C312S construct were treated with THZ1 or DMSO for 5 h with or without the presence of doxycycline. Cellular lysates were then probed for the indicated proteins or phosphoproteins. **d**, Overexpression of Flag-CDK7 C312 rescues the expression of a subset of transcripts in HeLa S3 cells. Log₂ fold change in gene expression in HeLa S3 cells expressing Flag-CDK7 wild type (x-axis) and Flag-CDK7 C312S (y-axis) after a 4 h treatment with 500 nM THZ1. **e**, GO molecular function analysis of transcripts increased by 1 log₂ order or more after expression of Flag-CDK7 C312S compared with Flag-CDK7 wild type in the presence of 500 nM THZ1.



Extended Data Figure 6 | THZ1 potently disrupts T-ALL proliferation.

a, THZ1, but not THZ1-R, exhibits strong antiproliferative effects against T-ALL cell lines. Cells were treated with THZ1, THZ1-R or DMSO vehicle for 72 h and assessed for antiproliferative effect by CellTiter Glo analysis. Error bars show \pm s.d. **b**, THZ1 causes cell-cycle arrest. Jurkat (top) and Loucy

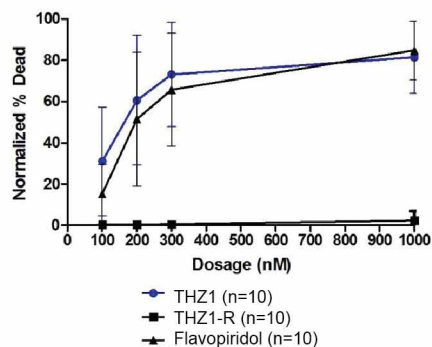
(bottom) T-ALL cells were treated with THZ1 for the indicated time periods. Cell-cycle progression was assessed using FACS cell-cycle analysis. 2N = G1, 4N = G2. **c**, Treatment with THZ1 decreases CDK1/2 T-loop phosphorylation. Jurkat cells were incubated with THZ1 for the indicated duration of time and lysates were probed for the specified proteins.



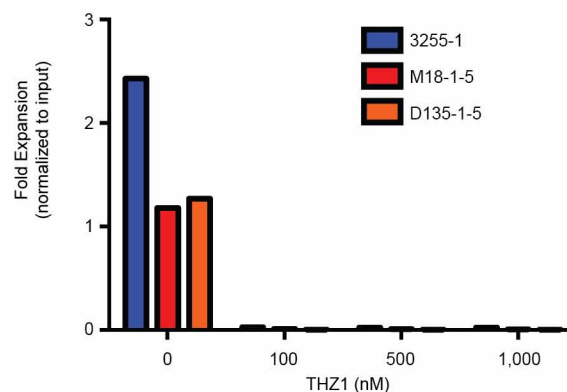
Extended Data Figure 7 | Treatment with THZ1 induces apoptosis in T-ALL cells. **a**, Representative annexin V and propidium iodide stainings for Jurkat cells incubated with THZ1 for the indicated amount of time and harvested to determine the percentage of apoptotic and/or dead cells by annexin V and propidium iodide staining, respectively. The percentage of cells in each cell population is shown in the four quadrants. **b**, Treatment with THZ1 induces apoptosis. Quantification of annexin V and propidium iodide staining data from **a**. Experiments were performed in biological triplicates. Error bars show \pm s.d. **c**, Representative annexin V and propidium iodide stainings for Loucy cells incubated with THZ1 for the indicated amount of time and

harvested to determine the percentage of apoptotic and/or dead cells by annexin V and propidium iodide staining, respectively. The percentage of cells in each cell population is shown in the four quadrants. **d**, Treatment with THZ1 induces apoptosis. Quantification of annexin V and propidium iodide staining data from **c**. Experiments were performed in biological triplicates. Error bars show \pm s.d. **e**, **f**, Sustained treatment with THZ1 induces apoptosis coincident with loss of RNAPII CTD phosphorylation and a reduction in anti-apoptotic proteins. Jurkat (**e**) and Loucy (**f**) cells were incubated with THZ1 for the indicated duration of time and lysates were probed for the specified proteins. Apoptosis was monitored by PARP cleavage.

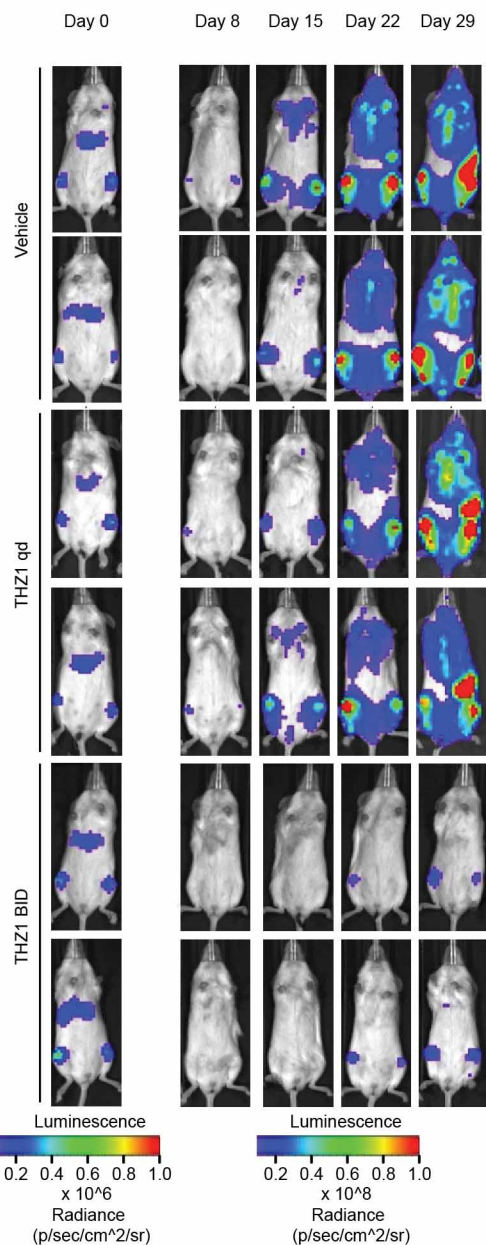
a



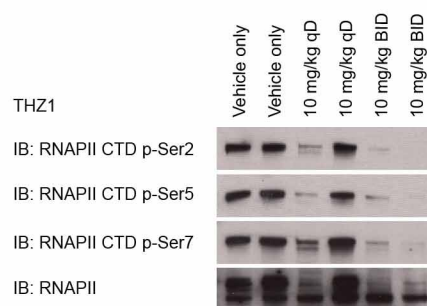
b



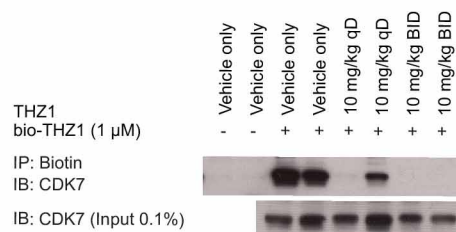
c



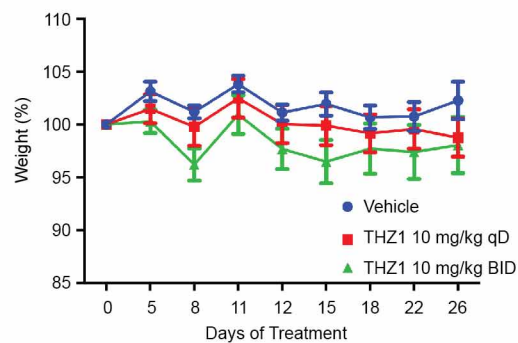
d



e



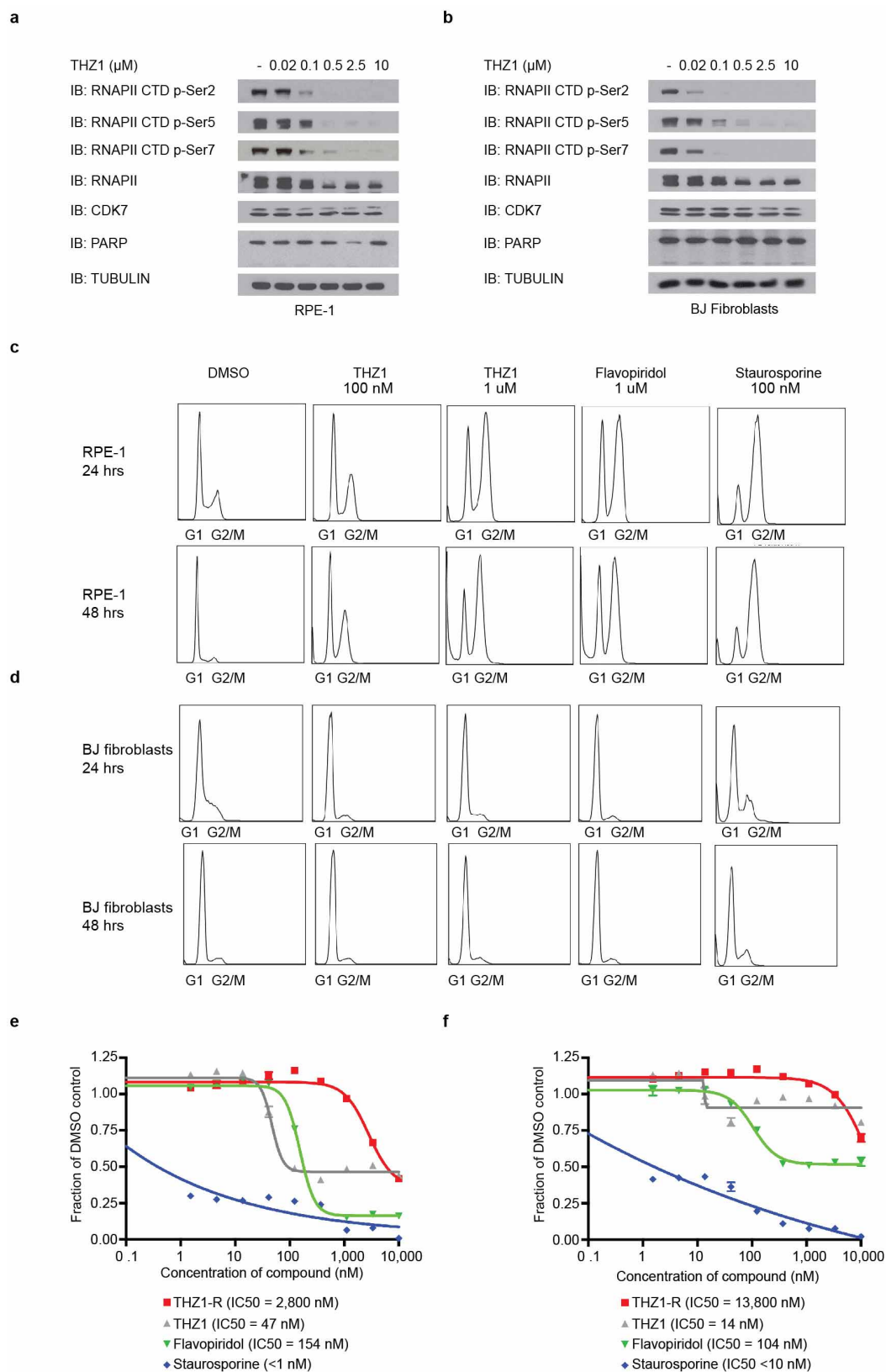
f



Extended Data Figure 8 | THZ1 demonstrates potent killing of primary chronic lymphocytic leukaemia cells and antiproliferative activity against primary T-ALL cells and *in vivo* against a human T-ALL xenograft.

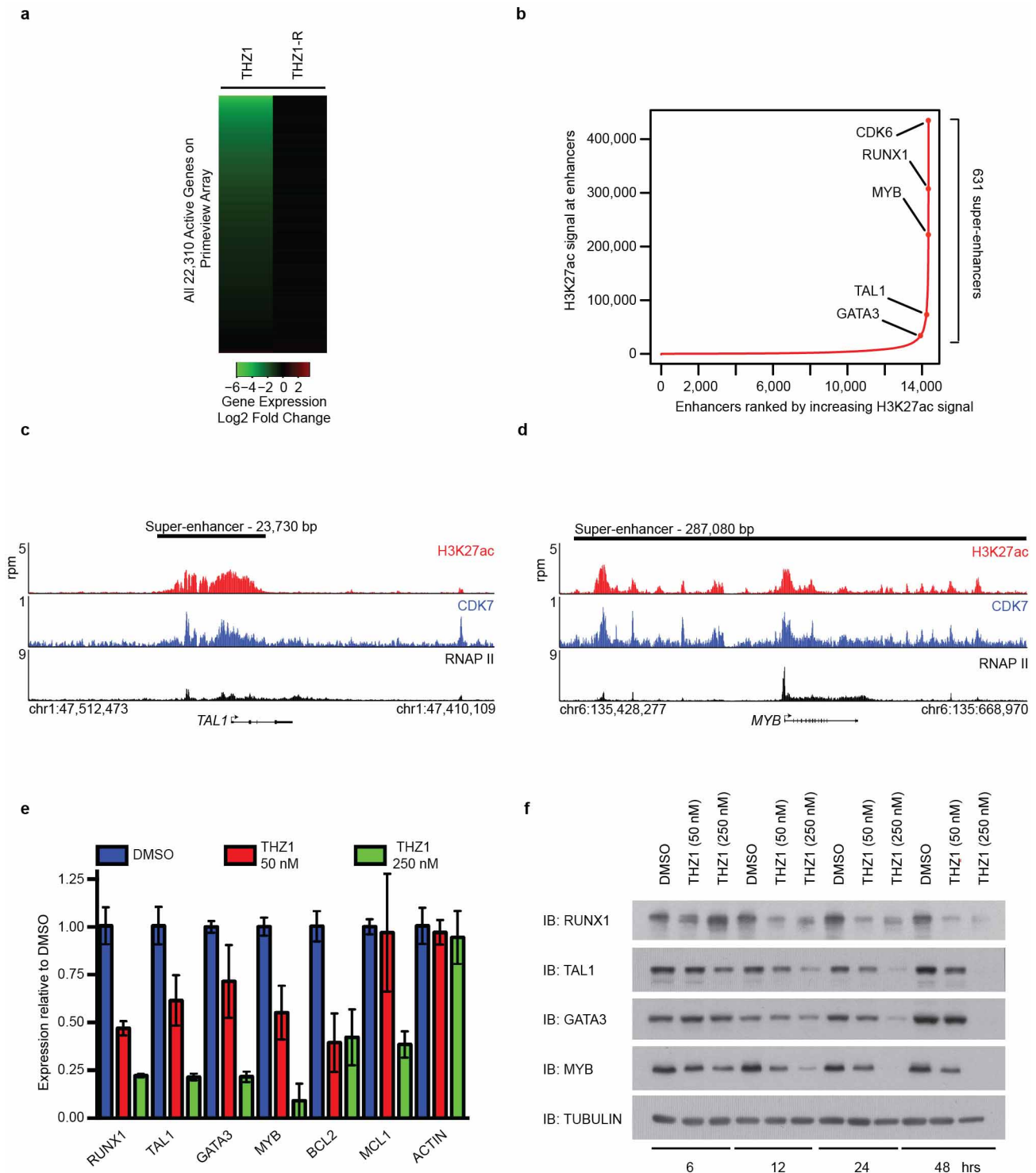
a, Patient-derived chronic lymphocytic leukaemia (CLL) samples were obtained and cultured *in vitro* for 24 hours in the presence of escalating doses of the specified compounds ($n = 10$ samples, 1 technical replicate per condition, per sample). Cell death upon compound exposure was evaluated using FITC Annexin V Apoptosis Kit I (BD Biosciences) and 1×10^4 events were collected and analysed using a BD FACSCanto II flow cytometer. Results shown are mean normalized percentage death based on Annexin V and propidium iodide single- and double-positive cells (\pm s.d.) normalized to baseline death in the vehicle (DMSO) control condition of each sample. Compounds tested were THZ1, THZ1-R and Flavopiridol (THZ1 versus THZ1-R $P = 1.5 \times 10^{-38}$; THZ1 versus Flavopiridol $P = 0.05$). P values were generated using an ANOVA model. **b**, Patient-derived xenografts (patient IDs 3255-1, M18-1-5 and D135-1-5; $n = 3$) were treated with THZ1 for 3 h followed by compound washout. An aliquot of input cells was then counted by flow cytometry using a known quantity of flow cytometry calibration beads (data not shown; Molecular Probes). The remaining cells were plated onto MS5-DL1 feeder cells in the presence of serum-free media (supplemented with $0.75 \mu\text{M}$ SR1, 10 ng ml^{-1} interleukin (IL)-7, 10 ng ml^{-1} IL-2). Seventy-two hours later, cultures were

harvested by vigorous pipetting with Trypsin, filtered through nylon mesh to deplete feeders, and counted by flow cytometry using a known quantity of flow cytometry calibration beads and with gating to discriminate between T-ALL cells and carryover feeders. The final cell number was normalized to the input cell number to calculate fold expansion. This experiment was performed once per patient-derived sample. **c**, Bioluminescent images of two representative mice treated with either vehicle control, 10 mg kg^{-1} THZ1 once daily (qD), or 10 mg kg^{-1} THZ1 twice daily (BID) for the indicated number of days. **d**, Spleen tissue from mice treated with THZ1 shows decreased RNAPII CTD phosphorylation. Mice were treated with THZ1 10 mg kg^{-1} once daily or twice daily or vehicle control. The animals were killed and spleen tissues were isolated. Lysates prepared from homogenized spleen tissue were probed for RNAPII CTD phosphopeptides. **e**, THZ1 binds directly to CDK7 in mouse tissues. Mice were treated with THZ1 10 mg kg^{-1} once daily or twice daily or vehicle control. The animals were killed and spleen tissues were isolated. Lysates prepared from homogenized spleen tissue were incubated with biotinylated THZ1 for 12 h at 4°C and 2 h at room temperature to induce covalent bond formation. Proteins pulled down were then probed for the presence of CDK7. **f**, Body weights of mice treated with either vehicle control, 10 mg kg^{-1} THZ1 once daily, or 10 mg kg^{-1} THZ1 twice daily over the duration of the drug treatment.



Extended Data Figure 9 | THZ1 inhibits RNAPII CTD phosphorylation and causes cell-cycle arrest in non-transformed cell lines. **a, b**, THZ1 inhibits RNAPII CTD phosphorylation. RPE-1 (**a**) and BJ fibroblasts (**b**) were treated with THZ1 or THZ1-R for 4 h. Cellular lysates were then probed with antibodies against the indicated proteins. **c, d**, THZ1 causes cell-cycle arrest in non-transformed cells. RPE-1 (**c**) and BJ fibroblast (**d**) cells were treated with

THZ1, Flavopiridol, Staurosporine or DMSO vehicle for the indicated time periods. Cell-cycle progression was analysed after permeabilization and staining with propidium iodide. **e, f**, THZ1 inhibits proliferation of non-transformed cell lines. RPE-1 (**e**) and BJ fibroblast (**f**) cells were treated with THZ1, THZ1-R, Flavopiridol or Staurosporine for 72 h and antiproliferative effect were determined by CellTiter Glo. Error bars show \pm s.d.



Extended Data Figure 10 | High-dose THZ1 reduces global steady-state mRNA levels, but low-dose THZ1 preferentially downregulates components of the *TAL1*/*RUNX1*/*GATA3* transcriptional circuitry. **a**, THZ1, but not THZ1-R, causes global downregulation of steady-state mRNA levels. Jurkat cells were treated with THZ1 (250 nM) or THZ1-R (250 nM) for 4 h. Total RNA was isolated and ERCC spike-in controls were added relative to cell number and analysed using Affymetrix PrimeView microarrays. Heatmaps displaying the log₂ fold change in gene expression versus DMSO for 22,310 genes expressed in DMSO conditions at 6 h in THZ1 or THZ1-R. **b**, Total H3K27ac ChIP-seq signal (length \times density) in enhancer regions for all stitched enhancers in Jurkat. Enhancers are ranked by increasing H3K27ac ChIP-seq signal. **c**, **d**, Gene tracks of H3K27ac (top), CDK7 (middle) and RNAPII (bottom) ChIP-seq occupancy at the TSS, gene body and enhancer

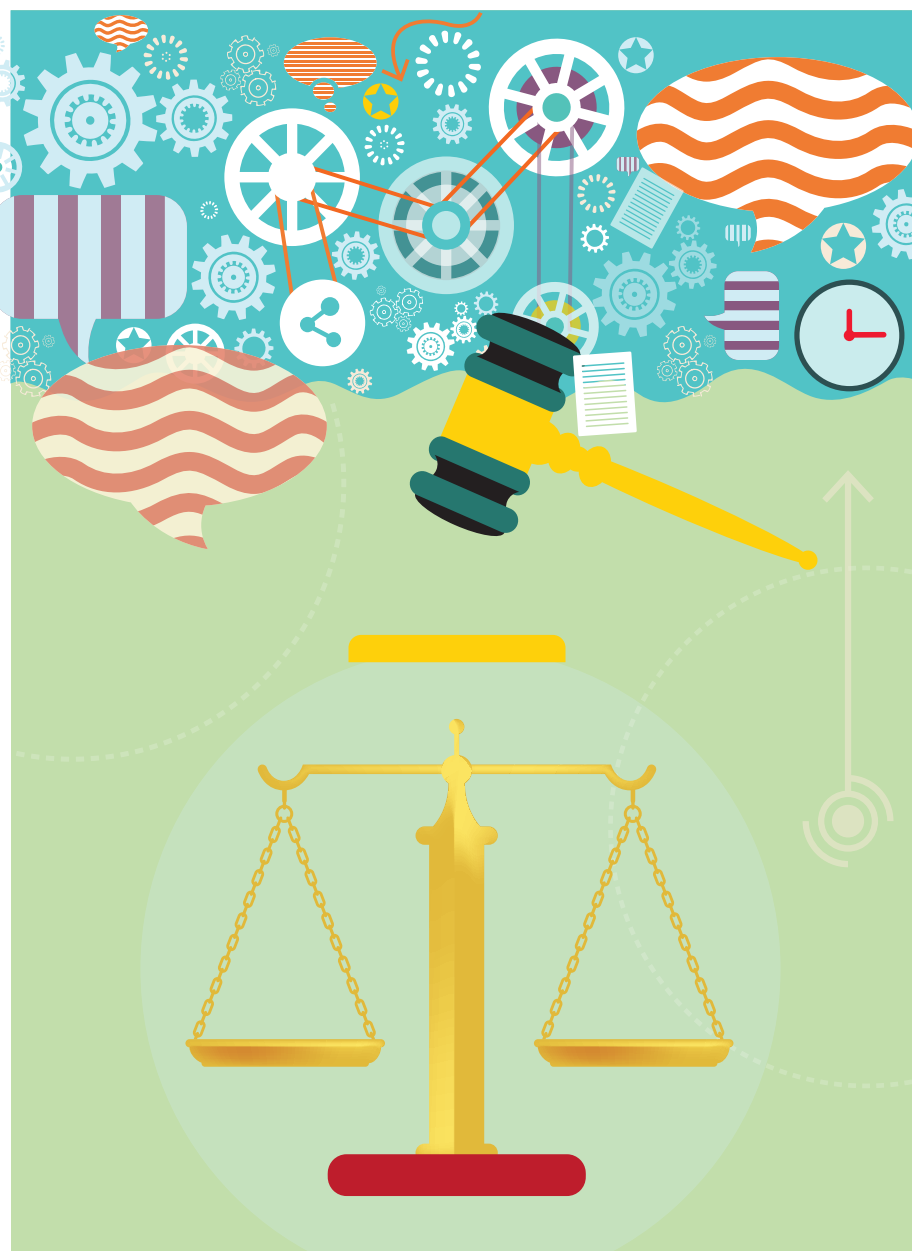
regions of *TAL1* (**c**) and *MYB* (**d**). **e**, THZ1 downregulates mRNA transcripts of the *TAL1*/*RUNX1*/*GATA3* transcriptional circuitry. Quantitative polymerase chain reaction with reverse transcription (RT-qPCR) expression analysis in Jurkat cells of transcripts identified as downregulated after THZ1 treatment. qPCR was carried out using Taqman probes according to the manufacturer's protocol. All experiments shown were performed in biological triplicate with each individual biological sample qPCR amplified in technical triplicate. Expression was normalized to *ACTB*, and fold change in expression was calculated relative to DMSO. Error bars show \pm s.d. **f**, THZ1 treatment reduces the protein levels of *TAL1*/*RUNX1*/*GATA3* transcriptional circuitry. Jurkat cells treated with THZ1 for the indicated time points were probed for the specified proteins.

CAREERS

EDUCATION Study probes how US graduate training can meet employers' needs **p.623**

FINANCES Report says money pressures are still squeezing US universities **p.623**

NATUREJOBS For the latest career listings and advice www.naturejobs.com



PATENT LAW

Finding a balance

Scientists who decide to pursue a legal career can enjoy fresh challenges while staying connected to the research world.

BY CAMERON WALKER

After Jason Rutt filled out a career-assessment survey as a teenager, he was given three job options: soldier, farmer or patent attorney. At the time, says Rutt, all he knew about patents was that the Nobel-prizewinning physicist Albert Einstein had worked in a patent office. It was not until years later, while pursuing a PhD in synthetic organic chemistry at the University of Nottingham, UK, that Rutt thought about patents again.

He spent six months of his PhD in the Nottingham laboratories of Boots, the international pharmacy and health-and-beauty chain. While there, he learned about the company's other departments, including the patent division, where attorneys secured patents for inventions and protected existing intellectual property from challenges. Rutt later returned to Boots to do research, but by then his desire to continue in research and development was waning. He decided to change tack: as a patent attorney, he realized, he would not only help clients to secure and protect intellectual property, but would also be able to work closely with researchers in many fields of science.

EASY SWITCH

The transition to law was fairly simple for Rutt, because he did not need a law degree, just on-the-job training and professional qualifications. By coincidence, a trainee position opened up in the patent department at Boots at the same time that Rutt decided he was interested in the field. He got the job. Now, as head of patents in the London office of the international intellectual-property firm Rouse, he works with pharmaceutical companies and start-up ventures in areas such as gene therapy and diagnostics.

Scientists move away from hands-on research and into the legal field for a variety of reasons. Some have a long-standing interest in law and policy; others believe that a career in law will allow them to be stronger advocates for research than they could be as scientists. Some consider a career in patent law because it dovetails with many scientific fields and will enable them to remain close to the research world. Patent attorneys may work directly with researchers to learn more about a client's techniques and inventions — they often hear about the latest ideas long before the research is published.

A patent attorney is not the only option for scientists who want to enter the legal field. ►

NIHAT DURSUN/ISTOCK VECTORS/GETTY

► Researchers with health or epidemiology backgrounds may choose to work as a legal counsel for public-health or regulatory agencies as the organizations apply for medical-research funding or examine bioethical issues. Graduates in environmental science or ecology may find positions helping government agencies to develop energy regulations or petitioning for endangered status for various species on behalf of environmental non-profit organizations (see 'From the lab to the law').

While studying the ecology of artificial reefs as part of a master's degree in environmental management, Margaret Peloso took classes in environmental law. That made her realize that she wanted to use her scientific background to develop policy, and she went on to pursue a PhD that combined science and policy while getting her law degree. She now focuses on environmental law and climate change in the Washington DC office of Vinson & Elkins, an international law firm specializing in energy and finance. Her work involves developing the firm's climate-change practice — she researches legal aspects of climate change and educates clients about potential risks and benefits, which fits in with her PhD on legal and policy issues surrounding sea-level rise.

WORLD CLASS

Qualifying as a legal professional varies from country to country. In the United Kingdom, prospective lawyers usually take an undergraduate degree in law. People who already have a degree — including one in science — can take a one-year postgraduate diploma in law. In both cases, extra coursework and training is required before being able to practise as a barrister or solicitor. But a law degree is not required for people with science and engineering backgrounds who want to become patent attorneys or to bring inventions to the European Patent Office (EPO). Instead, they do several years of on-the-job training, then sit national or EPO examinations.

In the United States, most of those wishing to become a practising attorney must gain an undergraduate degree and a law degree. After receiving their juris doctor (JD), they must then pass a state bar exam. (In a handful of states, it is possible to qualify to take the state bar exam after completing a lengthy legal apprenticeship.)

Yet the United States still has entry routes that do not require a law degree. Law firms often hire people with strong science backgrounds as technical specialists to help the firm to prepare patent applications for its technologies. These 'tech specs' often work directly with researchers and inventors to learn about their work, examine the scientific literature to find out whether similar techniques or ideas have already

FROM THE LAB TO THE LAW

A science background opens doors in the legal world



Shifting from a science-focused track to a legal career can be a bit bumpy, from the challenges of learning the language of law to studying for examinations. But it will be smoother for those who have a clear view of their target. Attorney Collette Adkins Giese in Minneapolis, Minnesota, had initially wanted to make an impact in the conservation world through research and teaching. After earning a master's degree in wildlife conservation and lecturing for several years, Giese realized that she wanted to be an advocate for environmental research.

So she returned to the University of Minnesota's Twin Cities campus for a joint juris doctor law degree and PhD programme in conservation biology. She spent her summer holidays as an intern at Earthjustice, an environmental-law organization, and for non-profit conservation groups, including Defenders of Wildlife in Washington DC. There, Giese made valuable connections and established her commitment to the legal and non-profit sectors — a double benefit because, she says, non-profits are often reluctant to hire newly minted lawyers. Law-school graduates do not always leave school with practical skills and could need on-the-job training, for which non-profit organizations may not have resources.

But after earning her law degree in 2005, she clerked for a judge who presided over criminal and contract issues, as well as some environmental law. She honed her writing skills by drafting the judge's orders and opinions. Then she spent several years at a large corporate law firm in Minneapolis

that had an established training programme for associates. With the firm, she worked on behalf of the plaintiffs in the 1989 *Exxon Valdez* oil spill and represented environmental non-profit organizations through the firm's pro bono programme. In her current role as a senior attorney for the Center for Biological Diversity in Tucson, Arizona, she focuses on protecting reptiles and amphibians.

For Giese, working as an attorney is the best way for her to use her scientific knowledge. "I really love the law, and I see how it changes and can make an impact. Important things get decided in courts," she says, "and I just want to be a part of that."

Another scientist-turned-lawyer is Heriberto Moreno. Moreno became interested in intellectual property as an undergraduate student at the University of Puerto Rico, where he had taken an elective class in technology transfer.

The class had given him the idea of training as a patent attorney. Towards the end of his microbiology PhD at the University of Virginia in Charlottesville, he perused the university's alumni database to find scientists-turned-patent attorneys and contacted them to learn more about their jobs. He also visited the school's career centre and looked online to find law firms that offer programmes and positions for technical specialists.

One that captured his interest was an intellectual-property firm based in Washington DC. Moreno contacted the head of the firm's biotechnology division and ended up working there as a technical specialist on patents in areas including nutraceuticals — products such as dietary and herbal supplements — pharmaceuticals and green technology. Over time, he took on more responsibility in drafting patent applications and realized that the next logical step was to become an attorney.

Moreno is now in his final year at Boston University's School of Law in Massachusetts. He also works as a technical specialist in areas including pharmaceuticals and biologics at the Boston office of the law firm McCarter & English. He has had to adapt to being older than most of his fellow students and managing the challenges of full-time study. But his mindset is different now that he has had a taste of his future career as a patent attorney. One of the best parts of patent law, he says, is its variety. "You just never know what you'll be working on." **C.W.**

been published and determine whether an innovation overlaps with technologies that have already been patented. Researchers who pass the US Patent and Trademark Office registration exam can then develop and file patent applications — although they cannot advise clients on legal issues or go to court if a patent is infringed.

If a scientist earns a JD and passes both patent and state bar exams, she or he can work in patent law and develop a full-service intellectual-property practice, which may involve working with clients on trademarks, copyrights and technology licensing. Scientist-attorneys can also help clients, particularly start-up firms, with issues such as entity formation, employment agreements and general legal services, and can practise in other fields of law as well. In some cases, employers will reimburse tuition fees or provide other forms of support to tech specs or patent agents who attend law school while working.

BACK TO SCHOOL

The idea of returning to higher education can be daunting for an early-career researcher, says Dianne Nicol, a law professor and deputy director of the Centre for Law and Genetics at the University of Tasmania in Hobart, Australia. But, she says, the extra training can help to develop one's career.

After earning a PhD in cell biology from Dalhousie University in Halifax, Canada, and a law degree in Tasmania, Nicol spent several years in a private legal practice, working on intellectual property as well as on contract work and personal-injury litigation. She now researches and writes about issues such as gene patenting, the privacy of genetic information, regulations underlying biobanking and direct-to-consumer genetic testing. Her research has been used to inform government reports on genetic privacy and health issues surrounding genetic patenting, and she was recently appointed to a three-member panel that reviews pharmaceutical patents. Nicol relies heavily on both her science background — from the fundamentals of genetics she learned as an undergraduate to her postgraduate training — and her legal education. “Even though it sounds like a long and laborious process, it's worthwhile,” she says.

The lingering effects of the global recession have meant that job prospects for fledgling US lawyers is gloomy: the American Bar Association announced this year that just 57% of the 46,776 people who received a JD in 2013 — the largest number of new JDs ever — had found long-term, full-time jobs in law that had required them to pass the bar examination. But the news is not all bad. People with science

backgrounds may have an edge on their peers, both in terms of law-school admission and in finding gainful employment after graduating, says Joy Baker Peacock, assistant director of the High Tech Law Institute at Santa Clara University School of Law in California. Many law schools and institutes are keen on training students who have physics, engineering and computer-science backgrounds — once qualified, such candidates may be appealing to employers in areas including semiconductors, photovoltaics and nanotechnology.

Peacock says that attorneys with a PhD in the life sciences are valuable in the biotechnology patent-prosecution field because they will have the necessary knowledge to work with clients and the officials on protecting intellectual property. Although litigators — attorneys who usually work for plaintiffs and defendants in patent-infringement cases — do not require a science background, she says, it can be an advantage. “Firms like to hire people that have a strong grounding in technical and scientific matters so that they will be able to get up to speed more quickly on the technical aspects of patent-infringement cases.”



“I really love the law. Important things get decided in courts. I just want to be a part of that.”

Collette Adkins Giese

in Tucson, Arizona. In 2012, Giese filed the largest-ever petition involving reptiles and amphibians to the US Fish and Wildlife Service to protect 53 species under the US Endangered Species Act. It was a 450-page document that involved nearly a year of preparation, including literature reviews and discussions with experts.

As a lawyer, she is now free to be an advocate for her interests. “You can't hide behind the notion that you're just presenting the facts anymore,” she says. “For me, that was such a relief.” ■

Cameron Walker is a freelance writer in Santa Barbara, California.

EDUCATION

Graduate skills survey

The Council of Graduate Schools in Washington DC is examining the professional-development requirements of PhD and master's students in science, technology, engineering and mathematics (STEM) programmes. The council will survey 500 member institutions and interview industry leaders to determine which skills are most important for STEM graduates and which remain unaddressed in US graduate programmes. Daniel Denecke, the council's associate vice-president for programmes, says that the study, which is funded by a US\$298,100 grant from the National Science Foundation, is focusing on industrial employers because they are the most likely to hire STEM graduates. Results will be available by summer 2016.

FUNDING

Marion Mason award

A US\$2.2-million bequest from the estate of a venerated US chemist will support early-career female chemists over the next 20 years. Recipients of the Marion Milligan Mason Award for Women in the Chemical Sciences will receive \$50,000, which may be used for laboratory supplies and equipment; publication costs; computer and technical support; and attendance at meetings. Applicants must have a tenure-track post at a US PhD-granting institution and must be US-born, naturalized citizens or permanent residents. The awards are administered by the American Association for the Advancement of Science in Washington DC. Applications are due by 15 September, and winners will be announced by May 2015.

FINANCIAL OUTLOOK

Continued squeezes

US universities are likely to face continued financial pressures over the next 12–18 months, says a report by Moody's Investors Service in New York. *Negative Outlook for US Higher Education Continues Even as Green Shoots of Stability Emerge* predicts that competition for tuition revenue, federal grants and state funding will affect regional public universities most; prominent private universities with large endowments will perform well. The negative outlook means that Moody's is more likely to give US universities poor credit ratings; as a result, they will incur higher borrowing costs and may have to cut back on hiring.

ONE OUT, ONE IN

The greatest gift.

BY AISLINN BATSTONE

Anitta and her granddaughter walked the avenue between snow-topped graves on their way to the church for Mass. Copenhagen's Assistens Cemetery seemed an appropriate place for a conversation like this, even if Karin was being obstinate.

"*Mormor*, there's no way I'll allow you to do it," Karin said.

"I want to. You're ready to become a mother."

"Mathias and I might adopt. If not, we'll wait in the queue. I have frozen eggs."

"Oh, a firstborn at age 70? That makes sense. Like me being as healthy as a 60-year-old."

"Good. You'll live for another 40 years."

"Not if it denies you children. I had three babies, you know." Christian, so fat with his dark hair and high colour. Nikolaj with his thoughtful eyes and the way he always wanted to be close to her. And Karin's mother Josefine, a noisy child whose shrieks made strangers smile.

"It's a different world, *Mormor*."

"Not a different world. Just different ways."

Somewhere nearby, the philosopher and theologian Søren Kierkegaard lay sleeping in the frozen earth. A scattering of snow decorated the evergreens and the bare branches. Through the trees they glimpsed the green copper spire of the Holy Cross Church.

"The only thing I wish," Anitta said, "is that I could see you with your child. I picture it all the time. You in the hospital bed with your baby in your arms."

"One day Mathias and I will have a baby. Maybe one who really needs us. Then you will have your wish."

Anitta shook her head. There were plenty of couples and not enough orphans. Pregnancy, birth, feeding a baby with her own body. Why should Karin miss out because her own generation was too selfish to die? "We won't talk about it anymore. But if I die, my 'one out, one in' goes to you." Such a request was legally binding, she'd made sure of that.

"Thank you," Karin said, clearly relieved that the conversation was over.

Karin felt sick to her stomach the day she had her contraceptive implant removed. They'd gone from 237,000th in the queue to top of the list in one leap.

To think that she and Mathias would now be actively trying to make a new person. A

new life that would start as a few cells in her womb and grow into an individual, a human being who would change her life.

She still felt shaky when she thought about what her grandmother had done for her. And now this nausea without even an embryo to cause it. Morning sickness. Mourning sickness. Traumatic to be handed the pot of a loved one's ashes and told that they had died so you could live a fuller life. A small comfort that the pot was her grandmother's favourite shade of blue. Had *Mormor* realized that Karin's life from now on would be joy and regret entwined like a helix of DNA?

A child! What would she do with a child? She only came across them every now and then, quiet little things surrounded by adults. Centenarians had taken over the playgrounds, whooping and swinging and sliding with their titanium hips and their fresh-grown organs.

"Ow!" Karin winced as the doctor injected local anaesthetic in her upper arm.

"Now hold still while I dig it out," the doctor said.

"So sweet. He looks like your uncle Nikolaj." Josefine, Karin's mother, curled her index finger into the newborn's hand. "Can I take care of him sometimes?"

"Of course, *Mor*. You're the third person who's asked. Actually, can you look after him on the night of our wedding anniversary? May 18th?"

"I'm sorry, Eric is taking me skiing that weekend."

Oh, that was just like her mother. But someone else would willingly babysit. "Help me feed him?"

"Here. I'll hold him while you get ready. What are you going to call him?"

"Maybe Nikolaj." Karin accepted the bundle from her mother. She brought his

head close to her breast and felt the strange sensation of a string twanging through the centre of

her body as the baby sucked and brought down her milk. "I wish *Mormor* could have met him."

"She looks so happy." Anitta touched Karin's cheek on the screen. It was wet, like her own.

She turned to face her guard. "The government has been extremely generous."

"It's time now," he said.

"I know." Last year they had interviewed her four times, asking her about the gold cross she wore around her neck and whether she was sure she wanted to take 'such a step'. *My relationship with God is personal. God wants what is best for me and my family.*

Anitta followed the guard out of the room and along the hospital corridor. They would give her the injection in the com-

fortable room she'd slept in since her 'death' over a year ago. They'd been so accommodating. She was truly grateful.

A nurse's warm hand covered hers, soothing the fear that fluttered like a moth in her chest. There'd been worse moments in her life, sickening fears, mostly involving her children and their accidents. In comparison, this was a question mark, equal parts hope and fear.

Death began with a beautiful dream.

A small girl playing in the snow. A young woman at university. A ringing of church bells. A mother with her first child.

Karin and her baby, dark-eyed Nikolaj. Anitta held him in her arms. She felt him grow chubby and tall. The smell of his hair. The softness of his skin.

The sound of the bells, snow on the grave-tops.

Love blossoming into unbearable ecstasy. Fireworks, noise.

Silence. Dark.

Peace.

Perfect peace. ■

Aislinn Batstone is no longer a practising Catholic, scientist or philosopher, but she is a practising mother. Always practising, never perfect, but the kids seem to be turning out OK.



JACEY

➔ **NATURE.COM**
Follow Futures:
@NatureFutures
f.gu.nature.com/mtoodm